

## Durham Research Online

---

### Deposited in DRO:

21 August 2014

### Version of attached file:

Published Version

### Peer-review status of attached file:

Peer-reviewed

### Citation for published item:

Hou, A. and Parker, L.C. and Wilman, D.J. and McGee, S.L. and Harris, W.E. and Connelly, J.L. and Balogh, M.L. and Mulchaey, J.S. and Bower, R.G. (2012) 'Substructure in the most massive GEEC groups : field-like populations in dynamically active groups.', *Monthly notices of the Royal Astronomical Society.*, 421 (4). pp. 3594-3611.

### Further information on publisher's website:

<http://dx.doi.org/10.1111/j.1365-2966.2012.20586.x>

### Publisher's copyright statement:

This article has been accepted for publication in *Monthly notices of the Royal Astronomical Society* © 2012 The Authors *Monthly Notices of the Royal Astronomical Society* © 2012 RAS Published by Oxford University Press on behalf of Royal Astronomical Society. All rights reserved.

### Additional information:

## Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

# Substructure in the most massive GEEC groups: field-like populations in dynamically active groups

Annie Hou,<sup>1\*</sup> Laura C. Parker,<sup>1</sup> David J. Wilman,<sup>2</sup> Sean L. McGee,<sup>3,4</sup>  
William E. Harris,<sup>1</sup> Jennifer L. Connelly,<sup>2</sup> Michael L. Balogh,<sup>4</sup> John S. Mulchaey<sup>5</sup>  
and Richard G. Bower<sup>3</sup>

<sup>1</sup>*Department of Physics & Astronomy, McMaster University, Hamilton ON L8S 4M1, Canada*

<sup>2</sup>*Max-Planck-Institut für Extraterrestrische Physik, Giessenbachstraße, D-85748 Garching, Germany*

<sup>3</sup>*Department Of Physics, University of Durham, Durham DH1 3LE*

<sup>4</sup>*Department of Physics and Astronomy, Univeristy of Waterloo, Waterloo, ON N2L 3G1, Canada*

<sup>5</sup>*Observatories of the Carnegie Institution, 813 Santa Barbara Street, Pasadena, CA 91101, USA*

Accepted 2012 January 17. Received 2012 January 16; in original form 2011 October 31

## ABSTRACT

The presence of substructure in galaxy groups and clusters is believed to be a sign of recent galaxy accretion and can be used to probe not only the assembly history of these structures, but also the evolution of their member galaxies. Using the Dressler–Shectman (DS) test, we study substructure in a sample of intermediate-redshift ( $z \sim 0.4$ ) galaxy groups from the Group Environment and Evolution Collaboration (GEEC) group catalogue. We find that four of the 15 rich GEEC groups, with an average velocity dispersion of  $\sim 525 \text{ km s}^{-1}$ , are identified as having significant substructure. The identified regions of localized substructure lie on the group outskirts and in some cases appear to be infalling. In a comparison of galaxy properties for the members of groups with and without substructure, we find that the groups with substructure have a significantly higher fraction of blue and star-forming galaxies and a parent colour distribution that resembles that of the field population rather than the overall group population. In addition, we observe correlations between the detection of substructure and other dynamical measures, such as velocity distributions and velocity dispersion profiles. Based on this analysis, we conclude that some galaxy groups contain significant substructure and that these groups have properties and galaxy populations that differ from groups with no detected substructure. These results indicate that the substructure galaxies, which lie preferentially on the group outskirts and could be infalling, do not exhibit signs of environmental effects, since little or no star formation quenching is observed in these systems.

**Key words:** galaxies: evolution – galaxies: groups: general – galaxies: statistics.

## 1 INTRODUCTION

The current theory of structure formation in the Universe is based on the standard  $\Lambda$  cold dark matter ( $\Lambda$ CDM) cosmological model, in which objects grow hierarchically from the initial matter density perturbations through mergers and accretion (e.g. Press & Schechter 1974; Lacey & Cole 1993; Springel et al. 2005). In order to test the theory of hierarchical structure formation, one must investigate the assembly history of the large structures in the Universe, namely galaxy groups and clusters.

A natural consequence of a hierarchical Universe is the existence of substructure within larger systems. Traditionally, substructure

has been defined as a kinematically distinct subhalo within a larger parent halo. A broader, more observationally motivated definition, and one that we will assume here, also includes separate haloes that are either merging to form a larger halo; gravitationally bound and infalling on to a pre-existing halo; or in the nearby large-scale structure, but not necessarily bound or infalling. In general, accreting structure does not have the same kinematic properties as the host, but whether or not substructure can be observed depends on how long it remains intact after infall. Early  $N$ -body simulations suggested that the assimilation of substructure into the host was rapid, providing no detectable long-lived features (Katz & White 1993; Summers, Davis & Evrard 1995). However, later work has shown that the lack of observable substructure in these simulations was due to poor resolution (Moore, Katz & Lake 1996). Indeed, high-resolution  $N$ -body simulations (e.g. Diemand et al. 2008; Springel

\*E-mail: houa2@physics.mcmaster.ca

et al. 2008) demonstrate that several hundred thousand subhaloes can exist in a Milky Way sized dark matter halo at a redshift of zero. Using semi-analytic models to study the substructure within individual galaxy to cluster-sized haloes, Taylor & Babul (2004) showed that accreting systems could survive within the host halo for many orbits, depending on the orbital parameters of the substructure upon infall.

These theoretical results suggest that substructure should be a detectable quantity, and numerical dark matter simulations of galaxy groups and clusters in a  $\Lambda$ CDM universe predict that approximately 30 per cent of all systems should contain substructure (Knebe & Müller 2000). Studies of individual clusters (e.g. Beers, Geller & Huchra 1982; Dressler & Shectman 1988; West & Bothun 1990; Bird 1994a; Colless & Dunn 1996; Burgett et al. 2004; Böhringer et al. 2010) indicate that a large fraction of systems show evidence of significant subclustering. The predicted theoretical value of 30 per cent is in agreement with some substructure studies of groups and clusters (e.g. Zabludoff & Mulchaey 1998b; Solanes, Salvador-Solé & González-Casado 1999), but several other results have demonstrated a much higher fraction of substructure; Bird (1994b) observed that 44 per cent<sup>1</sup> of their sample contained substructure, Dressler & Shectman (1988) observed 53 per cent and Ramella et al. (2007) find an extremely high substructure fraction of 73 per cent. Although the precise fraction of groups and clusters with substructure may still be a source of debate, the observed presence of any substructure strongly suggests that these systems grow in a hierarchical manner through the accretion of galaxies and smaller groups of galaxies.

Though galaxy clusters are amongst the largest structures in the local Universe, they do not represent the most common host environment for galaxies. Galaxy groups, which contain a few to tens of member galaxies, are the host environment of more than half of the present-day galaxy population (Geller & Huchra 1983; Eke et al. 2005). Despite the importance of groups in the build-up of large galaxy clusters, there have been few studies on the assembly history of groups themselves.

One method of probing group assembly histories is to look at the amount of substructure within these systems. The presence of such structure would indicate that the group has recently accreted galaxies (Lacey & Cole 1993). Studies have been carried out for galaxy groups in the local Universe by Zabludoff & Mulchaey (1998b), who observed significant [ $>99$  per cent confidence level (c.l.)] substructure in two of their six Local Groups. An interesting result of their analysis showed that the substructure was located on the outskirts of the systems, i.e.  $\sim 0.3\text{--}0.4 h^{-1} \text{ Mpc}$ , where  $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$ , from the core of the group. Based on these results, Zabludoff & Mulchaey (1998b) concluded that like rich galaxy clusters, some galaxy groups assembled in a hierarchical manner through the accretion of smaller structures from the field. A more recent study of local galaxy groups by Firth et al. (2006) found similar results, with roughly half of their sample showing significant substructure. Although, the findings of Zabludoff & Mulchaey (1998b) and Firth et al. (2006) provided an important first step in unveiling the assembly history of groups, their analysis was based on a small sample of very nearby systems. In order to gain a better understanding of the role of groups in the growth of structure, we must search for substructure in a larger sample of galaxy groups with highly complete spectroscopy that cover a wide range in red-

shift. Such studies have only become possible with recent deep spectroscopic surveys that have produced large group catalogues, such as the Sloan Digital Sky Survey (SDSS; Berlind et al. 2006; Yang et al. 2007), the Group Environment and Evolution Collaboration (GEEC; Carlberg et al. 2001; Wilman et al. 2005) and the higher redshift extension of GEEC (GEEC2; Balogh et al. 2011) optical group catalogues.

In addition to the role of the group environment in the growth of large-scale structure in the Universe, studies of substructure within groups allow us to probe galaxy evolution. Since groups have fewer members than galaxy clusters, any substructure present will have a stronger effect on the observed group properties (i.e. colours, blue or active fractions). If substructure traces accreting galaxies, one might expect a correlation between observed galaxy properties and substructure. Possible correlations could exist between substructure and the colours of galaxies in groups, or with substructure and star formation rates (SFRs). One of the main goals of this paper is to search for such correlations.

In this paper we search for substructure in a sample of intermediate-redshift galaxy groups from the GEEC group catalogue. In Section 2, we discuss the sensitivity and reliability of the Dressler–Shectman (DS) test for group-sized systems using Monte Carlo simulations. In Section 3, we apply the DS test to the GEEC groups and discuss the results of our analysis. In Section 4, we discuss the relationship between the presence of substructure and other indicators of dynamical state, such as the shape of the group velocity distribution. In Section 5, we look for correlations between substructure and the properties of member galaxies, such as colour and star formation rate, and discuss the possible implications of our findings. In Section 6, we present our conclusions. Additionally, we include an appendix (Appendix A), which provides detailed results of our Monte Carlo simulations.

Throughout this paper we assume a  $\Lambda$ CDM cosmology with  $\Omega_M = 0.3$ ,  $\Omega_\Lambda = 0.7$  and  $H_0 = 75 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

## 2 DETECTING SUBSTRUCTURE IN GROUPS

Numerous tests for substructure have been developed and carried out for cluster-sized systems (see Pinkney et al. 1996, for a thorough review). In a comparison of five 3D tests, which use both velocity and spatial information, Pinkney et al. (1996) determined that the DS test (Dressler & Shectman 1988) was the most sensitive test for substructure, for systems with as few as 30 members. In the following section we look at the DS test and determine its reliability and robustness for smaller group-sized systems.

### 2.1 The Dressler–Shectman (DS) test

Substructure manifests itself as detectable deviations in the spatial and/or velocity structure of a system. The aim of the DS test is to compute *local* mean velocity and velocity dispersion values, for each individual galaxy and its nearest neighbours, and then compare these to the global group values. Following the notation of Dressler & Shectman (1988), we define  $(\bar{v}, \sigma)$  as the mean velocity and velocity dispersion of the entire group, which is assumed to have  $n_{\text{members}}$  galaxies. Then for each galaxy  $i$  in the group, we select it plus a number of its nearest neighbours,  $N_{\text{nn}}$ , and compute their mean velocity  $\bar{v}_{\text{local}}^i$  and velocity dispersion  $\sigma_{\text{local}}^i$ . From these we compute the individual galaxy deviations,  $\delta_i$ , given as

$$\delta_i^2 = \left( \frac{N_{\text{nn}} + 1}{\sigma^2} \right) \left[ (\bar{v}_{\text{local}}^i - \bar{v})^2 + (\sigma_{\text{local}}^i - \sigma)^2 \right], \quad (1)$$

<sup>1</sup> 11/25 galaxies in their sample have significant ( $>95$  per cent) substructure based on results of the Dressler–Shectman statistics.

where  $1 \leq i \leq n_{\text{members}}$  (i.e.  $\delta_i$  is computed for each galaxy in the system). Dressler & Shectman (1988) originally developed the test for cluster-sized systems and the number of nearest neighbours used to compute the  $\delta_i$  values was relatively high (i.e.  $N_{\text{nn}} = 11$ ). Since substructure in galaxy groups is likely to have fewer than 11 constituent galaxies, we take  $N_{\text{nn}}$  to be  $\sqrt{n_{\text{members}}}$  following the methodology of authors who have previously applied the DS test to group-size systems (Silverman 1986; Pinkney et al. 1996; Zabludoff & Mulchaey 1998a). This ensures that large kinematic deviations of a few neighbouring galaxies do not get diluted by adding too many unassociated galaxies, and thereby lowering the computed  $\bar{v}_{\text{local}}^i$  and  $\sigma_{\text{local}}^i$  values in equation (1).

The statistic used in the DS test is the  $\Delta$ -value, given by

$$\Delta = \sum_i \delta_i. \quad (2)$$

A system is then considered to have substructure if  $\Delta/n_{\text{members}} > 1.0$  (Dressler & Shectman 1988). This method of using a threshold value to find substructure is referred to as the critical values method.

An alternative method of identifying substructure with the DS test is to use probabilities ( $P$ -values) rather than critical values. The  $P$ -values for the DS test are computed by comparing the observed  $\Delta$ -value to ‘shuffled’  $\Delta$ -values, which are computed by randomly shuffling the observed velocities and reassigning these values to the member positions, a procedure called ‘Monte Carlo (MC) shuffling’. The  $P$ -values are given by

$$P = \sum (\Delta_{\text{shuffled}} > \Delta_{\text{observed}}) / n_{\text{shuffle}}, \quad (3)$$

where  $\Delta_{\text{shuffled}}$  and  $\Delta_{\text{observed}}$  are both computed using equation (2) and  $n_{\text{shuffle}}$  is the number of ‘MC shuffles’, and therefore the number of  $\Delta_{\text{shuffled}}$  values, used to compute the probability. One can see from equation (3) that systems with significant substructure will have low  $P$ -values, since it is unlikely to obtain the observed  $\Delta$ -value randomly (equation 2).

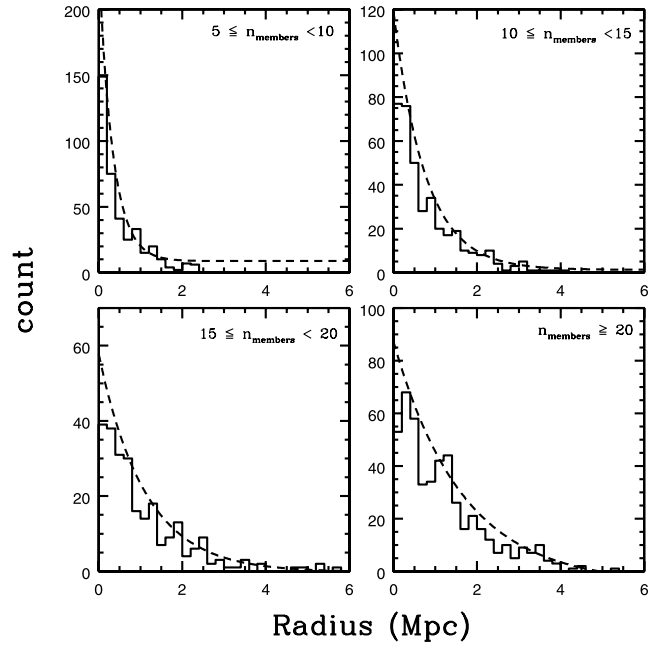
## 2.2 Monte Carlo simulations

Although Pinkney et al. (1996) have carried out an extensive investigation of the DS test, their analysis was performed on systems with a minimum of 30 members, and the majority of our intermediate-redshift groups have fewer member galaxies (see Section 3.1). Therefore, we perform our own Monte Carlo simulations in order to check the reliability, sensitivity and robustness of the DS test, using both the critical value and probability ( $P$ -value) methods, for group-sized systems. It should be noted that we specifically model our host mock groups after the observed GEEC group sample.

### 2.2.1 Generating the mock groups

Mock galaxy groups, both with and without substructure, are generated using Monte Carlo methods to assign member galaxy positions. These mock groups are then used to compute the false negative and positive rates of the DS test, and also determine the sensitivity of the test against a variety of input parameters.

The radial positions for the members of the mock groups are randomly drawn from fits to the projected group-centric radial distributions of the galaxies observed in the GEEC group catalogue (Wilman et al. 2005). The groups are divided into four bins based on the number of members in each group (Fig. 1), and fits to each bin are used to populate the mock groups. Since the GEEC groups span such a wide range of masses and group membership, we elect



**Figure 1.** Histograms of the radial distribution of the galaxies in the GEEC group catalogue stacked in bins of  $n_{\text{members}}$ . The dashed line corresponds to an exponential fit to the distribution, which we use to generate the radial positions of galaxies in our mock groups.

not to use a single fit to the radial distribution of all the group galaxies in our sample. Instead, we divide our sample into bins of group membership ( $5 \leq n_{\text{members}} < 10$ ,  $10 \leq n_{\text{members}} < 15$ ,  $15 \leq n_{\text{members}} < 20$  and  $n_{\text{members}} \geq 20$ ) and fit each distribution separately. It should be noted that we bin our sample by group membership, rather than mass or velocity dispersion, as we aim to study the false positive and negative rates of the DS test as a function of  $n_{\text{members}}$ . However, the results are similar if we bin by  $\sigma$  rather than  $n_{\text{members}}$ . The histograms and fits to the radial distributions of the binned groups are shown in Fig. 1. The general form of the radial distribution of member galaxies is given by

$$N \propto \exp(-\lambda R), \quad (4)$$

where  $R$  is the radial position of the given galaxy and  $\lambda = 2.98, 1.31, 0.902$  and  $0.606 \text{ Mpc}^{-1}$  for the  $5 \leq n_{\text{members}} < 10$ ,  $10 \leq n_{\text{members}} < 15$ ,  $15 \leq n_{\text{members}} < 20$  and  $n_{\text{members}} \geq 20$  bins, respectively. From these results it is clear that groups with fewer members have steeper radial distributions and smaller maximum group-centric radii (Fig. 1).

The redshifts of the member galaxies in the mock groups are randomly drawn from Gaussian distributions. The dispersions of the input Gaussians are taken to be the average velocity dispersion of the aforementioned group membership bins, which are the following: 300, 350, 400 and  $550 \text{ km s}^{-1}$ . These dispersion values are chosen for the mock groups in order to best mimic the systems in our GEEC sample. We also generate mock groups with the same input host velocity dispersion ( $\sigma_{\text{host}}$ ) for all values of  $n_{\text{members}}$  and find no significant difference in our results (see Appendix A for further detail).

As will be shown in Section 2.2.2, in order to determine the false negative rates for the DS test, we must include substructure within the host group. The positions and redshifts of the substructure

galaxies are drawn from a Gaussian distribution.<sup>2</sup> The free parameters in our mock groups with substructure are the velocity dispersion of the substructure ( $\sigma_{\text{redshift}}$ ), spatial dispersion of the substructure ( $\sigma_{\text{position}}$ ), location of the substructure, in both position ( $\epsilon_{\text{position}}$ ) and velocity ( $\epsilon_{\text{redshift}}$ ) space, the number of galaxies in the substructure ( $n_{\text{sub}}$ ), the number of galaxies in the host group ( $n_{\text{members}}$ ) and velocity dispersion of the host group ( $\sigma_{\text{host}}$ ). We briefly analyse the effects of each of these parameters on the robustness of the DS test in the following sections, and give a more detailed discussion in Appendix A.

### 2.2.2 Testing for false positives

Using the mock groups with no input substructure described in Section 2.2.1, we determine the false positive rates, or type I errors, which for the DS test occur when substructure is identified when none exists. The only parameter we vary for these systems is the number of member galaxies ( $n_{\text{members}}$ ).

We first compute the false positive rates using the critical values method. For each value of  $n_{\text{members}}$  we compute the  $\Delta$ -statistic (equations 1 and 2) and then classify a group as having substructure if  $\Delta/n_{\text{members}} > 1.0$ . With this criterion, we find that for all relevant values of  $n_{\text{members}}$  the false positive rates are extremely high. For example, for  $n_{\text{members}} = 20$ , the false positive rate is  $\sim 81$  per cent and even for larger systems,  $n_{\text{members}} = 100$ , we find that the rate is equally high. A similar result was observed by Knebe & Müller (2000) in a study of substructure in numerically simulated galaxy clusters. They found that for haloes with no substructure the  $\Delta/n_{\text{members}}$  values were often greater than 1.0, with values peaking closer to 1.4 for larger (i.e.  $n_{\text{members}} \sim 80$ –100) clusters. Although it is possible to better identify substructure in richer systems, such as clusters, with a higher value of  $\Delta/n_{\text{members}} \sim 1.4$ –1.6 (e.g. Knebe & Müller 2000; Girardi et al. 2005), we find that for group-sized systems this methodology could not simultaneously produce low false positive and false negative rates.

Alternatively, when  $P$ -values are used to identify substructure, we find that the false positive rates are much lower than with the critical values method and also remarkably stable for a wide range of group members (i.e.  $5 \leq n_{\text{members}} \leq 50$ ). From our mock groups we find that for significance levels of 0.01, 0.05 and 0.10, the false positive rates are 5, 10 and 15 per cent for all values of  $n_{\text{members}}$  tested. In other words, the DS test will falsely identify roughly 5 per cent more substructure than the desired significance level. Although these rates are higher than the expected values for a given significance level, they are still substantially lower than the rates obtained using the critical values method. Also, as long as one chooses significance levels of 0.01 or 0.05, the rate of false identifications is acceptably low. Based on these results, we rule out the critical values method and only perform the following analysis using the probabilities,  $P$ -values, method.

### 2.2.3 Testing for false negatives

Having determined the rate of false identifications obtained with the DS test, we now test for the rate of false negatives, or type II errors. For this statistic, the false negative rate measures how often the DS test fails to detect substructure when indeed it exists. The purpose

**Table 1.** Input parameter values for the ‘base’ mock groups described in Section 2.2.3, i.e. Monte Carlo groups with a zero false negative rate.

$n_{\text{members}}$	$n_{\text{sub}}$	$\sigma_{\text{host}}$ (km s <sup>−1</sup> )	$\sigma_{\text{position}}$ (Mpc)	$\epsilon_{\text{position}}$ (Mpc)	$\sigma_{\text{redshift}}$ (km s <sup>−1</sup> )	$\epsilon_{\text{redshift}}$ (km s <sup>−1</sup> )
10	4	350	0.01	0.5	100	1300
15	5	400	0.01	0.5	100	1300
20	5	550	0.01	0.5	100	1300
50	10	550	0.01	0.5	100	1300

of conducting these false negative tests is twofold: first, they help determine the reliability of the statistic; and secondly, by varying only one of the free input parameters (listed in Section 2.2.1) at a time, you can determine how sensitive the test is to each parameter. The latter places quantitative constraints on the maximum size and location of substructure that can be identified.

Before we begin looking at the input parameters, we first determine how reliable the test is at identifying ‘obvious’ substructure, that is to say galaxies that are tightly correlated and located far from the group centre, both in projection on the sky and along the line of sight (see Table 1 for input parameter values). For these mock groups, we find that substructure is correctly identified in almost all cases, using the  $P$ -value method and a significance level of 0.01, producing false negative rates of 0 or 1 per cent.

We then investigate the individual free parameters in more detail to determine how each alters the false negative rate. We briefly discuss the main results here and leave the detailed quantitative analysis, as well as full tables of false negative rates, for Appendix A.

Of the five free parameters, we find that the DS test is most sensitive to the number of galaxies in the substructure ( $n_{\text{sub}}$ ) and the location of the peak of the substructure’s velocity distribution with respect to the peak of the host’s ( $\epsilon_{\text{redshift}}$ ). Previous studies of substructure in galaxy clusters by Dressler & Shectman (1988) and Pinkney et al. (1996) showed that the DS test was unable to find substructure that was ‘superimposed’ with the highest density regions of the hosts, since the galaxies were spatially mixed. Pinkney et al. (1996) even stated that in these cases, any type of 3D test would not be able to accurately detect substructure. An important distinction, not made by these authors, is that two forms of ‘superposition’ can occur. Substructure can be superimposed with the host group either in projected angular position or in redshift space. Our analysis shows that the DS test is significantly more sensitive to  $\epsilon_{\text{redshift}}$  than to  $\epsilon_{\text{position}}$ . The test can usually identify substructure with superpositions as projected on the sky, but has a very difficult time with those along the line of sight. From equation (1), it is clear that only collections of neighbouring galaxies with a different local mean velocity and velocity dispersion will produce high  $\delta_i$  values, no matter their angular position.

In addition, we also find that the level of sensitivity of the DS test to  $\epsilon_{\text{redshift}}$ , the distance between the substructure’s and host’s peak velocity distribution, is dependent on the number of members in the host group. We find that for small groups (i.e.  $n_{\text{members}} < 20$ ), the input substructure needs to be further than  $2\sigma_{\text{host}}$  from the group centre in order to be detected. On the other hand, the DS test can identify substructure that is roughly  $2\sigma_{\text{host}}$  away from the peak of the host’s Gaussian velocity distribution in groups with more than 20 members. For even larger systems, i.e. clusters with  $n_{\text{members}} \geq 50$ , the input substructure can be located *within*  $1\sigma_{\text{host}}$  and still be identified.

<sup>2</sup> Tests with non-Gaussian input substructure have also been carried out, and for all reasonable distributions the results do not differ significantly from the results presented here.



Another result from this analysis is that the DS test is sensitive to the number of galaxies that are part of the substructure ( $n_{\text{sub}}$ ). Our simulations show that no matter the membership of the host group, the test cannot identify substructure with fewer than four members. Also, the minimum required number of members within the substructure increases with  $n_{\text{members}}$ . This is due to the fact that we set  $N_{\text{nn}}$  in equation (1) to  $\sqrt{n_{\text{members}}}$ . Therefore, more members in the host group means that the velocity information of more ‘neighbours’ will be used to calculate  $\bar{v}_{\text{local}}^i$  and  $\sigma_{\text{local}}^i$ . Thus, if there are too few galaxies in the substructure, their kinematic deviations can be washed out by other neighbouring host galaxies.

We find that for rich galaxy groups, with  $n_{\text{members}} \geq 20$ , the DS test can reliably identify true substructure, as has been shown by several other authors (e.g. Dressler & Shectman 1988; Pinkney et al. 1996; Knebe & Müller 2000). For systems with fewer member galaxies, such as poor groups with  $n_{\text{members}} < 20$ , the false negative rates are very low (<5 per cent) only for tightly correlated substructure galaxies with large kinematic deviations. In these poor groups, substructure that is more loosely associated with a velocity peak close to that of its host group is not as easily detected by the test.

Taking into account the results of both the false negative and positive tests, we conclude that for systems with  $n_{\text{members}} \geq 20$ , the DS test can reliably identify real substructure if the  $P$ -values method is employed using a c.l. of either 0.01 or 0.05. For groups with  $n_{\text{members}} < 20$ , we find that the percentage of groups with substructure identified by the DS test should be taken as a *lower limit*.

### 3 SUBSTRUCTURE IN THE GEEC GROUPS

#### 3.1 The GEEC group catalogue

The GEEC group catalogue is based on a set of  $\sim 200$  intermediate-redshift,  $0.1 < z < 0.6$ , galaxy groups initially identified in the second Canadian Network for Observational Cosmology (CNOC2) redshift survey (Yee et al. 2000; Carlberg et al. 2001). The CNOC2 survey observed  $\sim 4 \times 10^4$  galaxies covering four patches, totalling  $1.5 \text{ deg}^2$  in area, in the  $UBVR_{\text{CIC}}$  bands down to a limiting magnitude of  $R_{\text{C}} = 23.0$ . Spectra of more than 6000 galaxies were obtained with the multi-object spectrograph (MOS) on the Canada–France–Hawaii Telescope (CFHT), with 48 per cent completeness at  $R_{\text{C}} = 21.5$  (Yee et al. 2000). The GEEC group catalogue includes extensive follow-up spectroscopy with the Inamori Magellan Areal Camera and Spectrograph (IMACS; Connelly et al., submitted) and Low Dispersion Survey Spectrograph (LDSS2) on Magellan (Wilman et al. 2005), as well the FOcal Reducer and low dispersion Spectrograph (FORS2) on the Very Large Telescope (VLT; Connelly et al., submitted). We have also obtained multiwavelength imaging data, which include the following: X-ray observations with the *X-ray Multi-Mirror Mission–Newton (XMM–Newton)* and *Chandra X-ray Observatories* (Finoguenov et al. 2009), ultraviolet (UV) observations with *Galaxy Evolution Explorer (GALEX)*; McGee et al. 2011), optical observations with the Advanced Camera for Surveys (ACS) on the *Hubble Space Telescope (HST)*; Wilman et al. 2009), infrared observations with the Multi-band Imaging Photometer for *Spitzer* (MIPS) on the *Spitzer Space Telescope* (Tyler et al. 2011) and near-infrared observations with Isaac Newton Group Red Imaging Device (INGRID) on the William Herschel Telescope (Balogh et al. 2009), Infrared Array Camera (IRAC) on *Spitzer* (Wilman et al. 2008) and the Son of ISAAC (SOFI) on the New Technology Telescope (NTT; Balogh et al. 2009). In addition, improved opti-

cal imaging was obtained in the *ugrizBVRI* filters from the CFHT Megacam and CFH12K imagers (Balogh et al. 2009).

In addition to the follow-up observations of the CNOC2 fields, group membership has also been redefined by Wilman et al. (2005) using more relaxed algorithm parameters than those used by Carlberg et al. (2001). The original search parameters were optimized such that the group-finding algorithm would identify dense, virialized groups, while the Wilman et al. (2005) catalogue includes looser group populations that cover a wider range of dynamical states. For this reason, the GEEC group catalogue is ideal for the investigation of substructure within groups, since we are not restricted to dense group cores and are able to probe the surrounding large-scale structure.

#### 3.2 Analysis of the GEEC groups

We apply the DS test, as described in Section 2, to a subset of the GEEC groups. Although there are roughly 200 groups in the GEEC catalogue, the majority of systems have fewer than 10 members. In Section 2.2.3, we determined that in order to obtain a reliable measure of substructure, the minimum number of member galaxies for the DS test is  $n_{\text{members}} = 20$ , which leaves us with a subset of 15 groups. These groups are amongst the most massive GEEC groups with an average velocity dispersion of  $\sim 525 \text{ km s}^{-1}$ .

As previously mentioned, due to the relaxed membership allocation algorithm parameters, some of the GEEC groups have relatively large radial extents, and can be larger than the suggested maximum group virial radius ( $r_{200}$ ) of 1.0 Mpc (Mamon 2007), but for the purposes of detecting substructure we elect not to apply any radial cuts to our groups for our main analysis. Our definition of substructure is liberal, in that we include structure that may be infalling or structure that is in nearby large-scale structure but not necessarily bound to the host group. Thus, to ensure that we do not eliminate any possible detection of substructure, we analyse all galaxies identified as group members by the friends-of-friends algorithm applied in Wilman et al. (2005). Our decision to not apply radial cuts is further justified given that substructure is often in group and cluster outskirts (West & Bothun 1990; Zabludoff & Mulchaey 1998b).

For the distribution parameters we estimate  $\bar{v}_{\text{local}}^i$  and  $\bar{v}$ , from equation (1), as the group and local (i.e.  $i$ th galaxy and its  $\sqrt{n_{\text{members}}}$  nearest neighbours) canonical mean velocity, and  $\sigma_{\text{local}}^i$  and  $\sigma$ , also from equation (1), as the local and group intrinsic velocity dispersion, computed following the method outlined in Wilman et al. (2005). The dispersion uncertainties are computed using the jack-knife method (Efron 1982) and are given in Tables 2 and 3.

Of the 15 GEEC groups with  $n_{\text{members}} \geq 20$ , we find that four groups ( $\sim 27$  per cent) are identified as having substructure at the 99 per cent c.l. In Table 2, we list group properties,  $\Delta/n_{\text{members}}$  values, and  $P$ -values for the groups with substructure, identified as the systems that have  $P$ -values of less than 1 per cent. In Table 3 we list the same values, but for groups without substructure, i.e. systems that have  $P$ -values greater than 1 per cent. Also, it should be noted that although we list the  $\Delta/n_{\text{members}}$  critical values in Tables 2 and 3, we rely on the  $P$ -values to identify substructure in our groups. In addition, Tables 2 and 3 list the dynamical properties of the groups to be discussed in Sections 4.1 and 4.2. In Fig. 2, we plot both the intrinsic velocity dispersion of the group (top) and the total number of group members (bottom) versus the mean group redshift. Groups with and without substructure both span a wide range of velocity dispersions and group membership, indicating that there is no apparent redshift bias with regard to group velocity dispersions or  $n_{\text{members}}$  for the DS test.

**Table 2.** Group properties and DS statistics for the GEEC groups with substructure.

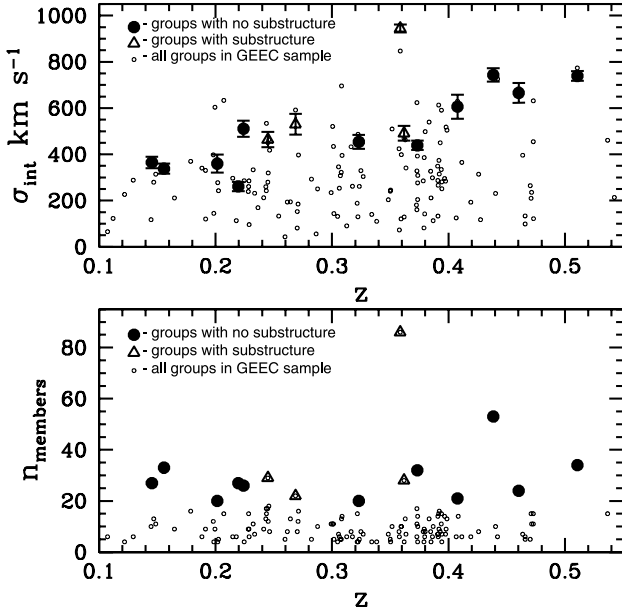
GEEC group ID	$n_{\text{members}}$	$z_{\text{group}}$	$\sigma$ ( $\text{km s}^{-1}$ )	$\Delta/n_{\text{members}}$	$P$ -value <sup>a</sup>	AD test classification	Shape of velocity dispersion profile
25	28	0.362	$491 \pm 32$	0.693	0	Non-Gaussian	Rising
208	22	0.269	$530 \pm 45$	0.841	0	Non-Gaussian	Rising
226	86	0.359	$944 \pm 17$	1.37	0	Non-Gaussian	Rising
320	29	0.245	$463 \pm 33$	0.751	0.003 83	Gaussian	Rising

<sup>a</sup>Using 100 000 ‘MC shuffles’. We identify groups with  $P$ -values  $< 0.01$  as having significant substructure.

**Table 3.** Group properties and DS statistics values for the GEEC groups with no substructure.

GEEC group ID	$n_{\text{members}}$	$z_{\text{group}}$	$\sigma$ ( $\text{km s}^{-1}$ )	$\Delta/n_{\text{members}}$	$P$ -value <sup>a</sup>	AD test classification	Shape of velocity dispersion profile
4	20	0.201	$360 \pm 39$	0.605	0.524	Non-Gaussian	Flat
38	34	0.511	$739 \pm 21$	0.910	0.0372	Non-Gaussian	Declining
104	27	0.145	$365 \pm 25$	0.659	0.0501	Gaussian	Flat
110	33	0.156	$338 \pm 22$	0.599	0.413	Gaussian	Declining
117	27	0.220	$261 \pm 20$	0.412	0.851	Gaussian	Flat
138	53	0.438	$743 \pm 29$	0.983	0.354	Non-Gaussian	Flat
238	21	0.408	$606 \pm 52$	0.826	0.135	Non-Gaussian	Flat
308	26	0.224	$511 \pm 35$	0.854	0.610	Gaussian	Declining
334	20	0.323	$454 \pm 30$	0.670	0.0563	Gaussian	Flat
346	32	0.373	$439 \pm 20$	0.612	0.232	Non-Gaussian	Flat
362	24	0.460	$666 \pm 43$	0.861	0.0817	Gaussian	Rising

<sup>a</sup>Using 100 000 ‘MC shuffles’. We only identify groups with  $P$ -values  $< 0.01$  as having significant substructure.



**Figure 2.** Top: intrinsic velocity dispersion  $\sigma_{\text{int}}$  versus the group redshift for the galaxy groups in our sample. Solid circles indicate groups with no substructure and open triangles indicate groups with identified substructure according to the DS test, at a 99 per cent c.l. The smaller open circles represent the dispersions of all the groups in the GEEC catalogue, the majority of which are not used in this analysis. Our  $n_{\text{members}} \geq 20$  GEEC sample tends to have higher velocity dispersions. Bottom:  $n_{\text{members}}$  versus the group redshift for the galaxy groups in our sample. Symbols are the same as the plot above.

Although the minimum membership cut needed to determine a reliable percentage of groups that contain substructure is  $n_{\text{members}} = 20$ , we can still apply the DS test to systems with fewer members and establish a lower limit on this fraction. Including groups with

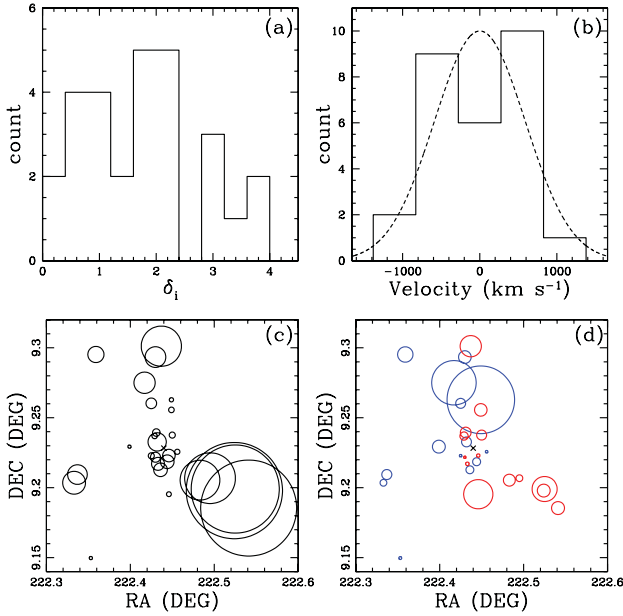
as few as 10 members increases our sample size to 63 systems. Applying the same methodology described above, we find that 11 of our 63 groups ( $\sim 17$  per cent) are identified as having significant ( $> 99$  per cent c.l.) substructure.

### 3.3 GEEC groups with substructure

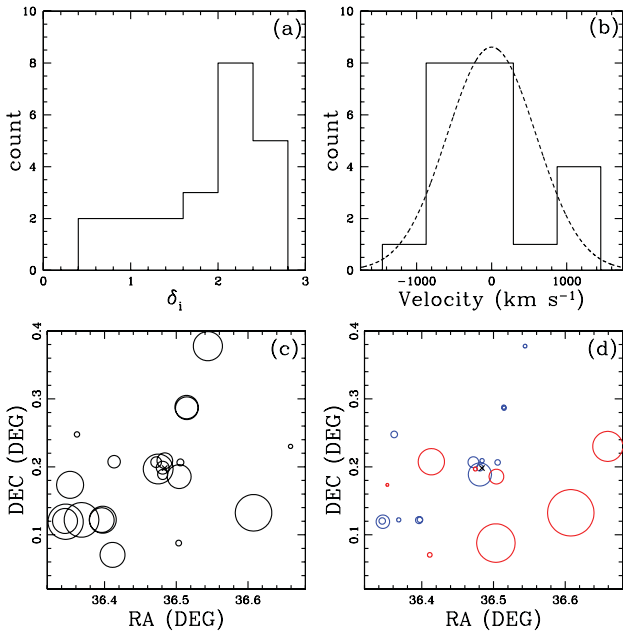
We examine the GEEC groups classified as having substructure by the DS test in detail to determine if we can identify the regions of localized substructure. By examining the available position and velocity information, we can find collections of galaxies that are kinematically distinct from the host group.

In the following analysis, we focus on small subgroups of galaxies that may be part of some localized substructure. This is done by looking at the  $\delta_i$  histogram (Figs 3a–6a), velocity histograms (Figs 3b–6b), ‘bubble-plots’ [i.e. position plots of the group members weighted by  $\exp(\delta_i)$ ] (Dressler & Shectman 1988) (Figs 3c–6c) and group-centric radial velocity-weighted (i.e.  $cZ_{\text{member}} - cZ_{\text{group}}$ ) position plots (Figs 3d–6d).

The  $\delta_i$  and velocity histograms provide an estimate on the amount of substructure and the dynamical state of the groups. The  $\delta_i$  histogram gives an overall view of the kinematic deviations, and the velocity histogram can be used to identify non-Gaussian features, such as multiple peaks. In order to look for local regions of substructure, one must look at both the ‘bubble-plot’ and velocity-weighted position plots in Figs (3)–(6). The ‘bubble-plots’ allow for the visual identification of candidate regions of local substructure. Since the size of the symbols in the ‘bubble-plot’ scales with a galaxy’s  $\delta_i$  value, large symbols correspond to strong local kinematic deviations from the global values. Thus, a collection of neighbouring galaxies with similarly large symbols, such as region A in Fig. 5(c), could indicate a kinematically distinct system. In order to confirm that these candidate regions are truly distinct, one must check that the candidate substructure galaxies also have similar velocities, since the sign or direction of the galaxy velocity is not taken into

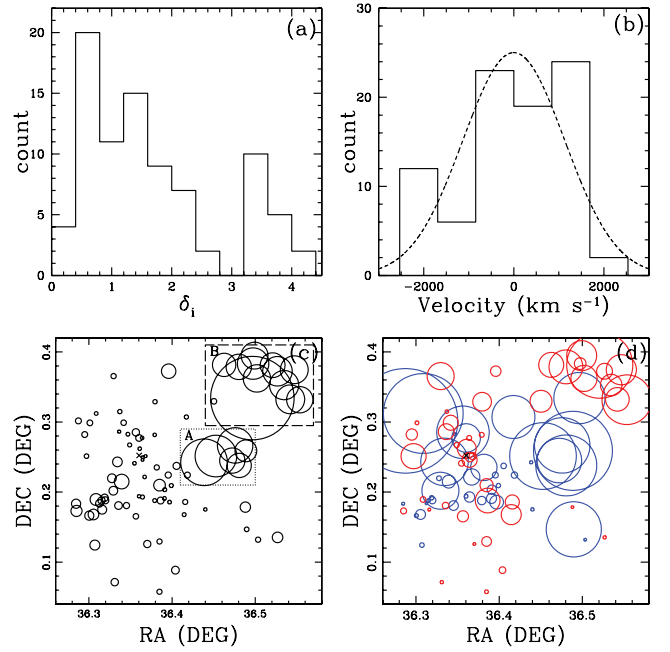


**Figure 3.** GEEC Group 25. (a)  $\delta_i$  histogram. (b) Histogram of the velocity distribution, where the dashed line indicates the best-fitting Gaussian velocity dispersion. (c) Dressler & Shectman (1988) ‘bubble-plot’ where the galaxy symbols scale with  $\exp(\delta_i)$ . (d) Position plot where the galaxy symbols scale with group-centric velocity [i.e.  $\exp(cz_{\text{member}} - cz_{\text{group}})/350$ ], blue symbols correspond to galaxies with negative group-centric velocities (i.e.  $cz_{\text{member}} - cz_{\text{group}} < 0$ ) and red symbols correspond to galaxies with positive group-centric velocities (i.e.  $cz_{\text{member}} - cz_{\text{group}} > 0$ ).

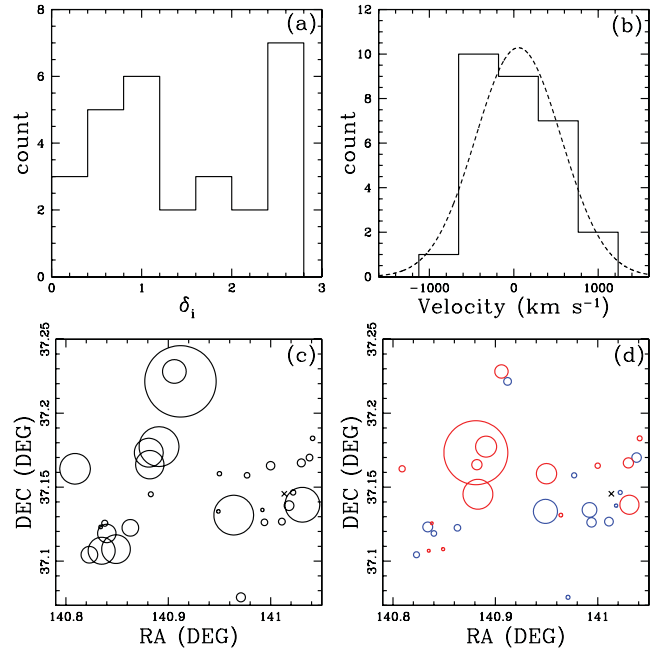


**Figure 4.** Same as Fig. 3 but for GEEC Group 208 and the velocity-weighted plot now scales as  $\exp(cz_{\text{member}} - cz_{\text{group}})/400$ .

account in the DS test (equation 1). To determine this, we look at the group-centric velocity-weighted position plot to see if the candidate local substructure galaxies have similar velocities. In Figs 3d–6d the red symbols correspond to positive group-centric velocities (i.e.  $cz_{\text{member}} - cz_{\text{group}} > 0$ ), blue symbols correspond to negative



**Figure 5.** Same as Fig. 3 but for GEEC Group 226 and the velocity-weighted plot now scales as  $\exp(cz_{\text{member}} - cz_{\text{group}})/600$ . The dotted box encompasses the first identified region of local substructure (region A) and the long-dashed box encompasses the second region of local substructure (region B).



**Figure 6.** Same as Fig. 3 but for GEEC Group 320.

group-centric velocities (i.e.  $cz_{\text{member}} - cz_{\text{group}} < 0$ ) and the symbol size scales with the magnitude of the velocity offset. We only identify galaxies as part of local substructure if neighbouring galaxies have similar large kinematic deviations, shown in the ‘bubble-plots’ and similar group-centric radial velocities, shown in the weighted position plots.

We will now discuss each of the four GEEC groups with substructure in detail. Using the methodology described above, we search



for candidate local regions of substructure in our sample. GEEC Group 25 (Fig. 3) has a collection of five galaxies, just south-east of the group centre, that all have high  $\delta_i$  values and comparable velocities. Similarly in GEEC Group 208 there are seven galaxies that lie south-west of the group centre, with equally high  $\delta_i$  values (Fig. 4c). However, when we look at these same galaxies in the velocity-weighted position plot (Fig. 4d), only five of these seven galaxies have comparable radial velocities. This result highlights the importance of looking at both the ‘bubble-plot’ and velocity-weighted position plots, as galaxies with large kinematic deviations may be correlated in position space, but not in velocity space.

In Fig. 5, we show the substructure analysis plots for GEEC Group 226 and from the ‘bubble-plot’ we see that there are two possibly distinct regions of localized substructure. The first region lies directly north-east of the group centre (region A in Fig. 5c) and includes the galaxy with the highest  $\delta_i$  value. All of the members within this given substructure have similar radial velocities (Fig. 5d). The second region of interest contains 11 galaxies that lie in the very north-east corner of the ‘bubble-plot’ (region B in Fig. 5c). Again, all of the galaxies within this particular substructure have similar group-centric velocities, though in the opposite direction of the members in region A.

From the ‘bubble-plot’ of GEEC Group 320 (Fig. 6c), we see two possible regions of substructure; one just north-west of centre and another further south-west of centre. The structure near the group centre does not appear to be very localized, as the velocities, though in the same direction, have significantly different magnitudes (Fig. 6d). The second region of high  $\delta_i$  values may actually be two separate structures. The velocity-weighted position plot in Fig. 6(d) reveals that three of the galaxies in the structure have similar negative group-centric velocities (i.e.  $CZ_{\text{member}} - CZ_{\text{group}} < 0$ ) and two have similar positive velocities (i.e.  $CZ_{\text{member}} - CZ_{\text{group}} > 0$ ). Therefore, we identify the collection of three galaxies, with the negative group-centric radial velocities, as the best candidate of substructure within this system.

In each of the GEEC groups with substructure, it appears that our identified regions of localized substructure lie on the outskirts or edges of the group, and may either be infalling on to a pre-existing system or in the nearby large-scale structure, but not necessarily bound or infalling. This result is in agreement with similar studies of substructure in Local Groups (Zabludoff & Mulchaey 1998b) and clusters (West & Bothun 1990).

It should also be noted that although the DS test has been shown to be reliable for group-sized systems, the number of members within the substructure can affect the results of the statistic. As discussed in Section 2.2, and shown in more detail in Appendix A, fewer (more) true members of substructure can increase (decrease) the rate of false negatives. However, we also show that the rate of false positives is consistently low for all of our mock groups (see Section 2.2). Thus, any detection of substructure in these systems is likely to be real.

### 3.4 Is the localized substructure gravitationally bound to the group?

The local regions of substructure we detect lie on the group outskirts and are possibly bound and infalling, or not bound, but close in large-scale structure. In order to help distinguish between these two possibilities, we apply a simple bound test to estimate whether or not the regions of localized substructure are gravitationally bound to the host group. This is done by computing the limits for bound systems, using a variation of the virial theorem as discussed in Beers

et al. (1982), and determining if the substructure falls within these limits. Beers et al. (1982) state that for a two-body system on a linear orbit, the Newtonian limit for gravitational binding, projected on to the sky is given by

$$V_r^2 R_p \leq 2GM \sin^2 \alpha \cos \alpha, \quad (5)$$

where

$$V_r = V \sin \alpha, \quad R_p = R \cos \alpha, \quad (6)$$

and where  $\alpha$  is the angle between the line joining the two-body system and the plane of the sky (see fig. 7 of Beers et al. 1982),  $M$  is the total mass of the entire system (substructure plus host group) and  $R$  and  $V$  are the true (3D) positional and velocity separations between the two objects.  $V_r$  is the line-of-sight relative velocity between the two bodies and  $R_p$  is the projected separation, both of which are measurable quantities.

The unknown quantity in equation (5) is the projection angle  $\alpha$ . Thus, to compute the limit between bound and unbound systems, one must work in  $\alpha$ - $V_r$  space to determine the probability that the system is gravitationally bound for any given projection angle. This is achieved by setting  $0^\circ \leq \alpha \leq 90^\circ$  and solving for  $V_r$  in equation (5), producing a distinct line in  $\alpha$ - $V_r$  space that clearly separates bound and unbound solutions (Fig. 7). One can then compute the probability of a bound solution, for a given projection angle, using the  $\alpha$ - $V_r$  plot.

To apply this methodology to our group sample, we treat our identified regions of local substructure as one body and the remaining galaxies as the second body, which we refer to as the ‘host’ group. The total mass of the system is taken to be the virial mass,  $M_{200}$  (i.e. the total mass within a radius that encloses a mean density of 200 times the critical density of the Universe at the redshift of the galaxy), of the GEEC groups as computed in Balogh et al. (2009) and given by

$$M_{200} = \frac{3^{3/2} \sigma^3}{G} \frac{1}{10H_0 (1+z)^{1.5}}, \quad (7)$$

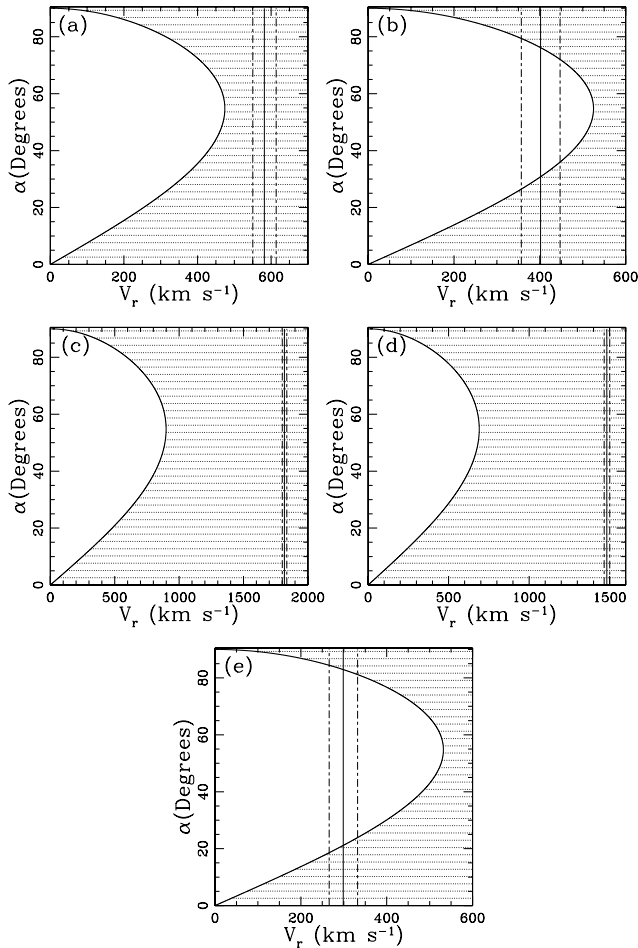
where  $\sigma$  is the measured intrinsic velocity dispersion. The measured  $V_r$  and  $R_p$  values are taken to be the distance between the  $R$ -band luminosity-weighted centres of the local substructure and the ‘host’ group centre, along the line of sight ( $V_r$ ) and projected on the sky ( $R_p$ ).

In Fig. 7, we plot the  $\alpha$ - $V_r$  plots for the five candidate regions of local substructure, as discussed in Section 3.3. The shaded regions indicate the areas spanned by unbound solutions, as given by equation (5), the solid black vertical line is the measured value of  $V_r$  and the long-dashed–short-dashed vertical lines indicate  $1\sigma$  deviations, taken to be the intrinsic velocity dispersion error.

Using the methodology described above and Fig. 7, we conclude that the identified local substructures in GEEC Groups 25 and 226 are not bound to the host group, while for GEEC Groups 208 and 320 the detected substructure is likely bound to the host group.

### 3.5 Substructure within 1 Mpc of the group centroid

In the previous section, we analysed a subset of 15 GEEC groups, with  $n_{\text{members}} \geq 20$ , without applying any radial cuts. Here we apply a 1.0-Mpc cut, which is the suggested maximum virial radius for groups (Mamon 2007), on the same subset of groups and reapply the DS test. Applying this radial cut, while still requiring a minimum number of 20 member galaxies, reduces our sample from 15 to 5 groups (GEEC Groups 110, 138, 226, 308 and 346). With this radial cut, we find that all five groups are identified as not having



**Figure 7.**  $\alpha$ – $V_r$  plot for GEEC Groups (a) 25, (b) 208, (c) region A and the host group of 226, (d) region B and the host group of 226 (d) and (e) 320. The unshaded regions correspond to bound solutions and shaded regions correspond to unbound solutions of the virial theorem given by equation (5). The solid line corresponds to the measured value of  $V_r$  (i.e. the line-of-sight velocity difference between the substructure and host group centres). The dashed lines correspond to  $1\sigma$  deviations, taken to be the errors on the intrinsic velocity dispersion.

**Table 4.** Groups properties and DS statistics values for the GEEC groups with a 1.0-Mpc radius cut.

GEEC group ID	$n_{\text{members}}^a$	$\bar{v}$ ( $\text{km s}^{-1}$ )	$\sigma_{\text{int}}$ ( $\text{km s}^{-1}$ )	$P$ -value
110	26	−15.3	350	0.403
138	23	−226	730	0.792
226	25	−174	847	0.110
308	25	2.52	512	0.507
346	26	−80.4	434	0.128

<sup>a</sup>Group membership after the 1.0-Mpc radius cut.

substructure (see Table 4) according to the DS test. The only group that was previously identified as having substructure, prior to the 1.0-Mpc cut, is GEEC Group 226. In Fig. 5, it is evident that although there are galaxies with relatively high  $\delta_i$  values close to the group centre, the members with the highest  $\delta_i$  values lie near the edge of the group. In fact, the most significant feature in the ‘bubble-plot’ (Fig. 5) is on the top-right corner of the plot, far from the group centre.

If we include all of the groups with  $n_{\text{members}} \geq 10$  after a 1.0-Mpc radial cut, we find that two of 33 groups ( $\sim 6$  per cent) of our sample contain significant substructure. Again, we note that for systems with fewer than 20 members, the DS test can only provide a lower limit on the amount of substructure present.

#### 4 CORRELATIONS BETWEEN SUBSTRUCTURE AND OTHER INDICATORS OF DYNAMICAL STATE

Having identified the GEEC groups that contain substructure, we can also look at other dynamical properties, i.e. the velocity distributions and velocity dispersion profiles (VDPs), to determine if there are any correlations with substructure.

##### 4.1 Comparison with the dynamical classification of velocity distribution

If a correlation does exist between substructure and recent galaxy accretion, one would expect that the groups with substructure should also be dynamically complex, with perhaps non-Gaussian velocity distributions. In Hou et al. (2009), we established a classification scheme to distinguish between dynamically relaxed and complex groups. Using the Anderson–Darling (AD) goodness-of-fit test, we are able to determine how much a system’s velocity distribution deviates from Gaussian. This is done by comparing the cumulative distribution function of the ordered data, which in our case is the observed velocity distribution, to the model Gaussian empirical distribution function using computing formulae given in D’Agostino & Stephens (1986). The AD statistic is then converted into a probability, or  $P$ -value, using results determined via Monte Carlo methods in Nelson (1998). A system is then considered to have a non-Gaussian velocity distribution if its computed  $P$ -value is less than 0.01 corresponding to a 99 per cent c.l.

We now apply this scheme to our sample of 15 GEEC groups, with  $n_{\text{members}} \geq 20$  and no radial cut, to compare the dynamical state with the detection of substructure.<sup>3</sup> The results of our dynamical classification scheme indicate that eight of the 15 groups are classified as having non-Gaussian velocity distributions, at the 99 per cent c.l., and are thus dynamically complex (see Tables 2 and 3). Of the four GEEC groups with substructure, only GEEC Group 320 shows a velocity distribution consistent with being Gaussian. Five groups, GEEC Groups 4, 38, 138, 238 and 346, are identified as being dynamically complex, but do not contain any substructure according to the DS test.

Pinkney et al. (1996) found that different statistical tools (i.e. 1D, 2D and 3D tests) probe the dynamical state of a system at varying epochs. Using  $N$ -body simulations, these authors determined that 1D tests, such as the AD test, are most sensitive to scenarios when substructure passes through the core of the host group (i.e. core crossing). During this time, substructure can become spatially mixed within the host group, and if the substructure is loosely bound, then it may be difficult to detect with 3D tests, such as the DS test. Thus, groups with non-Gaussian velocity distributions may contain substructure that is missed by the DS test.

<sup>3</sup> The dynamical classifications in this paper differ from those in Hou et al. (2009). The reason for this difference is that in Hou et al. (2009) we applied a 1.0-Mpc radial cut to the GEEC groups, whereas no radial cut is applied in this analysis.

## 4.2 Comparison with the velocity dispersion profiles

In a study of the VDPs of galaxy clusters, Menci & Fusco-Femiano (1996) found a correlation between the efficiency of merger activity in the cluster core and the shape of the VDP, where actively interacting systems had strongly rising profiles. We presented a similar correlation for galaxy groups in Hou et al. (2009), where we found that groups classified as dynamically complex (i.e. non-Gaussian velocity distributions) also had rising VDPs. We now look at the VDPs of the current GEEC group sample to see if there is a similar relationship between the shape of the profile and the detection of substructure.

The VDPs are computed following the method outlined in Bergond et al. (2006). Unlike traditional methods of computing binned projected velocity dispersions, this technique generates a smoothed VDP. This is done by using a ‘moving window’ prescription, which takes into account the contribution of every radial velocity measurement at each computed radius. The values are binned with an exponentially weighted moving window given by

$$w_i(R) = \frac{1}{\sigma_R} \exp\left[-\frac{(R - R_i)^2}{2\sigma_R^2}\right], \quad (8)$$

where  $\sigma_R$  is the width of the window, which can be constant or a function of radius  $R$ , and  $R_i$  are the radial positions of the members of the system. The projected velocity dispersions are then defined as

$$\sigma_p(R) = \sqrt{\frac{\sum_i w_i(R)(x_i - \bar{x})^2}{\sum_i w_i(R)}}, \quad (9)$$

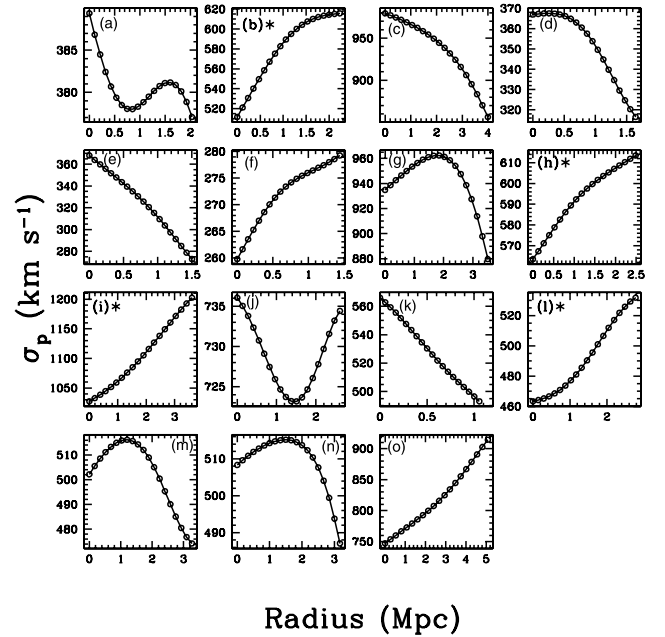
where the  $x_i$  are the radial velocities and  $\bar{x}$  is the mean velocity of the system.

We compute VDPs for our GEEC group sample and the profiles are shown in Fig. 8 using a window width,  $\sigma_R$ , equal to one-third the maximum group radius. It should be noted that the projected velocity dispersions do not include any redshift or instrumental error corrections, so the intrinsic dispersion values (Tables 2 and 3) are generally lower than the projected velocity dispersions.

Comparing the profiles of the groups with and without substructure, we find that all four GEEC groups identified as having substructure (Groups 25, 208, 226 and 320) also have strongly rising profiles. In contrast, almost all of the groups with no detected substructure, with the exception of GEEC Group 362, either have flat or generally decreasing VDPs, within the intrinsic velocity dispersion error. See Tables 2 and 3 for a description of the shape of the VDP for each GEEC group. Thus, we do indeed observe a correlation between detectable substructure and rising VDPs. As previously mentioned, studies of rich galaxy clusters suggest that a strongly increasing profile may be a signature of merger activity or galaxy interactions (Menci & Fusco-Femiano 1996), but an alternative explanation of rising VDPs is the presence of subclumps with different mean velocities (Girardi et al. 1996; Barrena et al. 2007). In our case, it is likely that the increasing profile is being caused by the kinematically distinct substructure, which has a different mean velocity from the host group.

## 5 CORRELATIONS BETWEEN SUBSTRUCTURE AND GALAXY PROPERTIES

Thus far, we have looked at possible correlations between substructure and other indicators of the dynamical state of a system. We now compare properties of the member galaxies to see if there are any differences between the galaxies in groups with and without substructure.



**Figure 8.** VDPs for the GEEC groups with  $n \geq 20$ : (a) Group 4; (b) Group 25; (c) Group 38; (d) Group 104; (e) Group 110; (f) Group 117; (g) Group 138; (h) Group 208; (i) Group 226; (j) Group 238; (k) Group 308; (l) Group 320; (m) Group 334; (n) Group 346; and (o) Group 362. The plots with asterisks in the top-left corner indicate groups that have been identified as having significant substructure. The intrinsic velocity dispersions, with errors, for each group can be found in Tables 2 and 3.

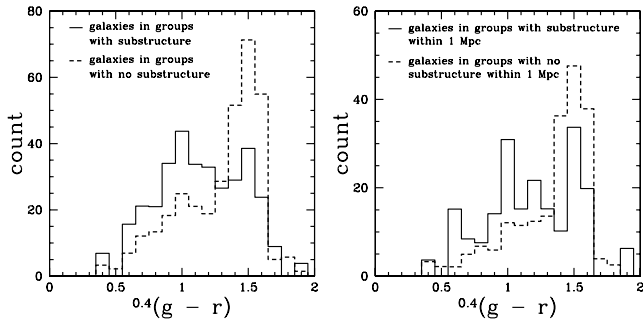
### 5.1 Substructure and colour

In the following analysis we compare the  $^{0.4}(g-r)$  colours, which have been corrected for galactic extinction and  $k$ -corrected to a redshift of  $z = 0.4$  (Balogh et al. 2009), and blue fractions,  $f_b$ , for the groups with and without substructure.

In addition to extinction and  $k$ -corrections, we also apply a completeness correction to address the differing spectroscopic coverage. We apply magnitude weights that depend on whether the group had follow-up spectroscopy. The weights we apply are similar to those derived in Wilman et al. (2005), except we do not include any radial weights. We compute weights in  $r$ -band magnitude bins of 0.25, and up to a limit of  $r = 22.0$ , which is the limit of the unbiased Magellan spectroscopy. These weights are then applied to all of the member galaxies in our sample.

In Fig. 9 we show the completeness weighted  $^{0.4}(g-r)$  histograms for the galaxies, with  $r < 22.0$ , in groups with substructure (solid line) and for those in groups with no detected substructure (dashed line). From this figure, it is clear that both histograms are bimodal with a distinct red sequence and blue cloud. Although both colour distributions have the expected bimodal shape, it is obvious that galaxies in the groups with and without substructure come from very different parent colour distributions. The groups with no identified substructure have a well-populated red sequence, while the groups with substructure appear to have a much more dominant blue cloud. Results from a two-sample Kolmogorov–Smirnov (KS) test on the unbinned  $^{0.4}(g-r)$  distributions show that these two samples are distinct at the 99 per cent c.l.

We also plot colour distributions of the groups with and without substructure with a 1.0-Mpc radial cut applied to all groups in our sample (Fig. 9, right). Although, the blue cloud for the groups with substructure is less populated when a radial cut is applied, it is



**Figure 9.** Left:  $^{0.4}(g-r)$  colour histograms for the GEEC groups with substructure (solid line), which has been normalized to match the number count of the groups with no substructure and for the GEEC groups with no detectable substructure (dashed line). Both histograms have  $r$ -band magnitude-based weights which have been computed to take into account the differing spectroscopic coverage between the original CNOC2 survey and the follow-up Magellan survey. Right: same as the figure on the left except a 1.0-Mpc radial cut has been applied to all groups in the sample.

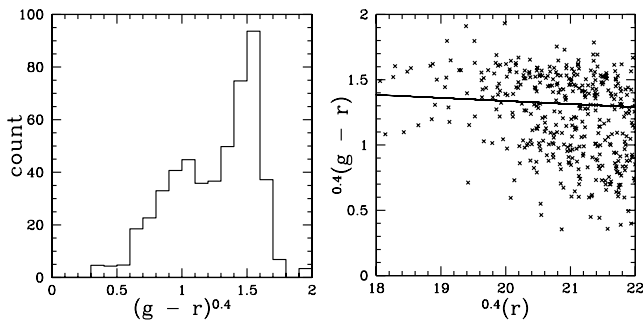
still more populated than the groups without substructure and the two distributions are still statistically distinct. This result indicates that the increase in blue galaxies in the groups with substructure is coming not only from galaxies at radii greater than 1.0 Mpc. Our findings are similar to those of Ribeiro, Lopes & Trevisan (2010) who showed that there are many more red galaxies in dynamically evolved group systems, even out to 4 virial radii.

To obtain a more quantitative comparison, we compute the fraction of blue galaxies within each sample. We take a multistage approach to determine the appropriate colour cut needed to distinguish between the red sequence and blue cloud. First, we apply an initial colour cut of  $^{0.4}(g-r) = 1.2$ , based on the minimum value in the colour distribution of all the member galaxies in our sample, shown in Fig. 10. We then determine a linear fit to all of the galaxies with  $^{0.4}(g-r) > 1.2$  and set the colour cut to be one standard deviation below the fit to the red sequence. The final colour cut is determined to be

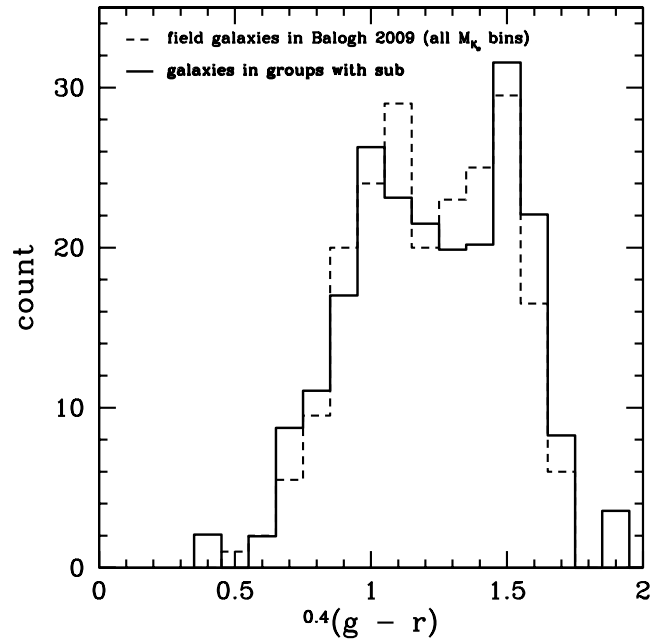
$$^{0.4}(g-r) = -0.0236(^{0.4}r) + 1.810, \quad (10)$$

and is represented by the solid line in the colour-magnitude diagram shown in Fig. 10.

The blue fraction,  $f_b$ , is then computed as the ratio of galaxies with  $^{0.4}(g-r)$  values that fall below equation (10) to the total number of



**Figure 10.** Left: weighted  $^{0.4}(g-r)$  histogram of all the member galaxies in our group sample. Right:  $^{0.4}(g-r)$  versus  $r$  colour-magnitude diagram of all the galaxies in our sample, with no completeness correction. The solid line indicates the colour cut used to distinguish between the red sequence and the blue cloud.



**Figure 11.**  $^{0.4}(g-r)$  colour histograms for the GEEC groups with substructure (solid line) and for all of field galaxies in the GEEC groups found in fig. 6 of Balogh et al. (2009) (dashed line). It should be noted that the colour distribution for the field galaxies is a sum of all the counts in each magnitude bin of fig. 6 of Balogh et al. (2009), since our group galaxy sample covers the entire magnitude range. In addition, the group sample has been normalized to match the number count of the field sample.

galaxies in the sample. The error in the blue fraction is computed using confidence intervals (CIs) derived from the beta distribution (Cameron 2010). This method has been shown to more accurately determine CIs, especially for small samples, over traditional methods such as Poisson errors, which systematically underestimates the width of the CIs. We find that  $f_b = 41 \pm 3$  per cent for the groups without substructure and  $f_b = 69 \pm 4$  per cent for the groups with substructure. Thus, the groups with substructure have a significantly higher blue fraction, which is clear in the  $^{0.4}(g-r)$  histograms of Fig. 9. We note that although these blue fractions are derived from weighted colour distributions, they are in agreement with the unweighted values,<sup>4</sup> suggesting that the applied magnitude weights do not affect the observed differences in colour.

In addition, we also compare the blue fractions of the individual galaxies identified as being part of the local substructure (see Section 3.3) to the other members of the group. We find that for the galaxies in the identified regions of localized substructure,  $f_b \simeq 74 \pm 8$  per cent, and for the galaxies not in the substructure,  $f_b \simeq 50 \pm 3$  per cent. This suggests that the observed increase in the blue fraction of groups with substructure is being enhanced by the galaxies in the identified regions of local substructure.

We now compare our groups with substructure sample to a sample of intermediate-redshift field galaxies. In Fig. 11 we reproduce the  $^{0.4}(g-r)$  histogram for the field galaxies in fig. 6 of Balogh et al. (2009) (dashed line), summing up all of the counts in each of the

<sup>4</sup>  $f_b = 44 \pm 4$  per cent for the groups without substructure and  $f_b = 68 \pm 5$  per cent for the groups with substructure.



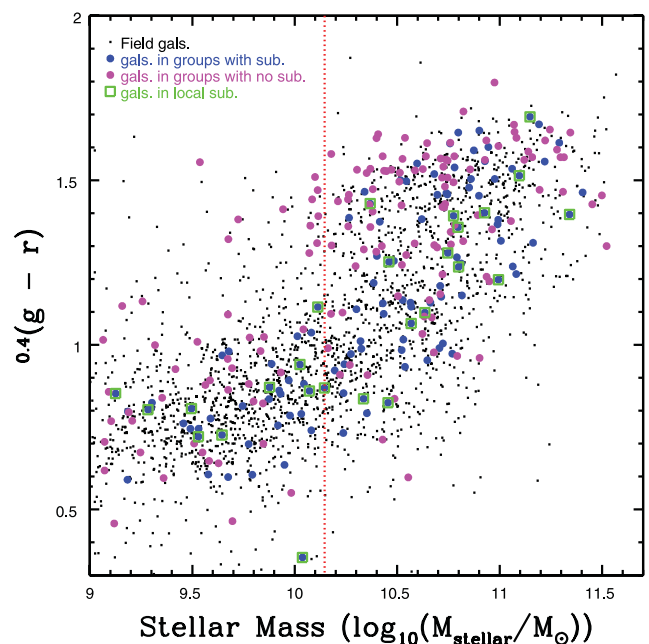
quoted  $M_{K_s}$  bins in order to get the total colour distribution.<sup>5</sup> We also overplot the colour histogram for the galaxies in groups with substructure, except we now apply an  $M_i$  magnitude cut in order to match the  $M_{K_s}$  range used in Balogh et al. (2009). Both the field galaxies in the Balogh et al. (2009) sample and the galaxies in our groups with substructure lack a prominent red sequence and have well-populated blue clouds. Despite subtle differences in the two colour histograms, a two-sample KS test shows that the  $^{0.4}(g-r)$  colours of the field galaxies and groups with substructure galaxies very likely come from the same parent distribution ( $P$ -value = 0.67).

Our blue fractions can be compared to those computed in the zCOSMOS spectroscopic survey (Lilly et al. 2007), where Iovino et al. (2010) determined the blue fractions of isolated and group galaxies at various redshifts in their sample. At a redshift of  $z = 0.4$ , Iovino et al. (2010) found  $f_b \sim 70$  per cent for isolated galaxies and  $f_b \sim 45$  per cent for group galaxies.<sup>6</sup> From these results it is clear that our observed blue fraction of 69 per cent for the galaxies in groups with substructure is significantly higher than the observed zCOSMOS group sample, but is in agreement with their field sample. On other hand, the  $f_b$  values for our GEEC groups with no substructure roughly agree with the zCOSMOS group sample.

## 5.2 Substructure as a function of colour, stellar mass and star formation rates

In the previous sections we compared the  $^{0.4}(g-r)$  colours of the galaxies in groups with substructure to those in groups with no detected substructure. Here we again compare the colours of the member galaxies but now as a function of stellar mass and specific star formation rate (SSFR) (defined as the ratio of the SFR to the stellar mass). Additionally, we compare the SSFR distributions of the galaxies in groups with and without substructure. It should be noted that since only three of the four original CNOC2 fields were targeted with *GALEX*, the sample used in the following analysis contains fewer galaxies than that used in the colour analysis of Section 4.3. In our sample of 15 ( $n_{\text{members}} \geq 20$ ) GEEC groups, 275 group galaxies have measured SFRs, stellar masses and colours, while 401 group members have well-determined colours.

The stellar masses and SFRs for the GEEC sample were obtained from spectral energy distribution (SED) template fitting to all of the available photometry. Detailed discussion of the methodology is presented in McGee et al. (2011), but we give a brief summary here. The photometry used in the SED-fitting process typically included photometry in the  $K$ ,  $i$ ,  $r$ ,  $g$ ,  $u$ , NUV and FUV bands (see Balogh et al. 2009, for details of the GEEC photometry). The observed photometry was then compared to a large grid of model SEDs constructed using the Bruzual & Charlot (2003) stellar population synthesis code and assuming a Chabrier initial mass function. Following the methodology of Salim et al. (2007), McGee et al. (2011) created a grid of models that uniformly sampled the allowed parameters of formation time, galaxy metallicity and the two-component dust model of Charlot & Fall (2000). The star formation history was modelled as an exponentially declining base rate with random bursts of star formation of varying duration and relative strength. The model magnitudes were obtained by convolving these model



**Figure 12.**  $^{0.4}(g-r)$  versus stellar mass for field galaxies (black dots), galaxies in groups with substructure (blue circles), galaxies in groups with no substructure (magenta circles) and galaxies identified as part of localized substructure (green squares). The red dotted line indicates the stellar mass limit ( $M_{\text{stellar}} = 1.4 \times 10^{10} M_{\odot}$ ) at median redshift ( $z \sim 0.3$ ) of the GEEC sample.

SEDs with the observed photometric bandpasses at nine redshifts between 0.25 and 0.60.  $\chi^2$ -minimization was then performed by summing over all of the models and taking into account the observed uncertainty on each point. The  $1\sigma$  uncertainties in stellar mass, when compared to both mock groups and other independent estimates, are of the order of 0.15 dex and the SFRs have been averaged over the last 100 Myr (McGee et al. 2011). It should be noted that there may be additional systematic uncertainties due to, for example, the initial mass function assumed in the fitting procedure.

In Fig. 12 we show  $^{0.4}(g-r)$  versus stellar mass for the field galaxies (black dots), galaxies in groups with substructure (closed blue circles), galaxies in groups with no identified substructure (closed magenta circles) and the galaxies identified as being part of local substructure (open green squares). The dotted red vertical line in Fig. 12 represents the stellar mass limit of  $1.4 \times 10^{10} M_{\odot}$  at the median redshift ( $z \sim 0.3$ ) in the GEEC group sample (McGee et al. 2011). It should be noted that this stellar mass limit shifts to lower (higher) masses for groups at lower (higher) redshifts. From Fig. 12, it is clear that the galaxies in groups with substructure lie preferentially along the blue cloud, as discussed in Section 4.3, but we also see that these galaxies span a similar mass range as the galaxies in groups with no identified substructure. A two-sample KS test of the stellar mass distributions of the galaxies with and without substructure shows that the two distributions likely come from the same parent distribution ( $P$ -value = 0.55). The similar stellar mass distribution of the two populations tells us that the DS test does not simply detect substructure in groups which are probed further down the stellar mass function.

Now we examine the SSFRs for groups with and without substructure. Following McGee et al. (2011), we define actively star-forming galaxies to have  $\log_{10}(\text{SSFR}) > -11$  and the fraction of actively star-forming galaxies to be  $f_{\text{active}} =$

<sup>5</sup> It should be noted that although only a fraction of our sample actually have measured  $M_{K_s}$  values, those that do span the entire range of magnitudes quoted in fig. 6 of Balogh et al. (2009).

<sup>6</sup> We look at Sample II of Iovino et al. (2010), as this is the data set that corresponds best to our GEEC sample.

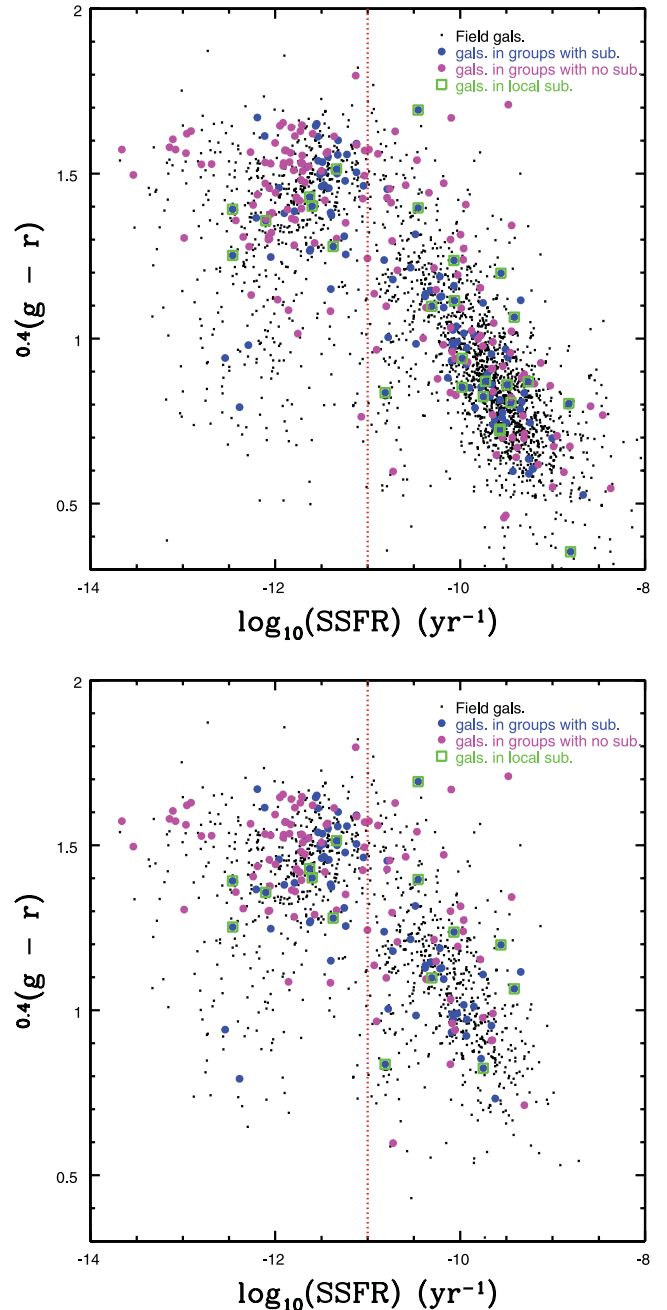


$n_{\text{actively star-forming galaxies}}/n_{\text{total}}$ . We find that for galaxies in groups with substructure,  $f_{\text{active}} = 63 \pm 8$  per cent, and for groups with no detected substructure,  $f_{\text{active}} = 49 \pm 5$  per cent. These active fractions agree with the blue fractions found in Section 5.1.

Both blue and active fractions are used as independent indicators of quiescent versus actively star-forming galaxies. However, colour and SSFR probe significantly different time-scales, and might be telling us something different about the star formation history of galaxies. For instance, a dust-enshrouded star-forming galaxy would be classified as ‘red’, and therefore quiescent, based on colour but would be classified as actively star-forming based on SSFR. A better approach is to look at colour and SSFR simultaneously (Weinmann et al. 2006). We follow this approach in Fig. 13 where we plot  $^{0.4}(g-r)$  versus SSFR for all the galaxies in our sample (top) and with a  $M_{\text{stellar}} > 1.4 \times 10^{10} M_{\odot}$  cut applied to the sample (bottom). The colour scheme is the same as in Fig. 12, except the red dotted line now corresponds to the division between actively star-forming and quiescent/passive galaxies. From Fig. 13, we can see that there is a correlation between colour and SSFR with two well-populated regions of the plot that correspond to ‘red and passive’ and ‘blue and active’ galaxies, where active refers strictly to actively star-forming galaxies. In Table 5, we list the percentage of all galaxies in our sample that populate each region of the colour–SSFR space in Fig. 13, with errors computed using the methodology described in Cameron (2010). Similarly, Table 6 lists the same information but for galaxies above the stellar mass completeness limit of  $1.4 \times 10^{10} M_{\odot}$ . From these tables we see that the galaxies in groups with substructure have significantly more blue and actively star-forming galaxies than groups with no substructure, though slightly less than the fraction observed in the field. This result indicates that environmentally driven mechanisms of star formation quenching are not as efficient in groups with observed substructure. We note that although the percentages within each region of colour–SSFR space differ between the whole versus stellar mass limited sample, the general trends remain the same. Since the mass-selected sample only includes galaxies with the highest stellar masses, we expect a decrease in the ‘blue and active’ region due to stellar mass trends (i.e. higher mass galaxies are preferentially more red and passive) (Iovino et al. 2010; Peng et al. 2010).

While the majority of galaxies are either ‘red and passive’ or ‘blue and active’, a non-negligible fraction appear to lie in the other two regions of Fig. 13. Even if we take into account the small uncertainties in colour (typically 0.02 mag) and the uncertainties in SSFR (McGee et al. 2011 quote  $1\sigma$  errors of the order of 0.25 dex), a measurable fraction of galaxies still remain in these two regions. From Table 5, we see that  $\sim 9$  per cent of all galaxies in the field and group samples reside in the ‘blue and passive’ region. This value is higher than the  $\sim 1.1$  per cent observed by Weinmann et al. (2006), though it appears similar to the results of Lara-López et al. (2010). Note that both these results are based on SDSS galaxies. Although we do observe a substantial population of ‘blue and passive’ galaxies in Fig. 13, we acknowledge that obtaining accurate measures of SSFR for galaxies with low SFRs is notoriously difficult. As discussed in McGee et al. (2011), the errors in the measured SSFRs increase for galaxies with lower values, and some of these galaxies may have underestimated SSFRs.

The final region of the  $^{0.4}(g-r)$ –SSFR plot in Fig. 13 corresponds to ‘red and active’ galaxies. While this population is negligible for the groups with substructure sample and small for the field sample (see Table 5), it contains  $\sim 11 \pm 4$  per cent of the galaxies in the groups with no detected substructure and  $\sim 14 \pm 4$  for the stellar mass limited sample. Although this population may be a result of



**Figure 13.** Top:  $^{0.4}(g-r)$  versus SSFR for field galaxies (black dots), galaxies in groups with substructure (blue circles), galaxies in groups with no substructure (magenta circles) and galaxies identified as part of localized substructure (green squares). The red dotted line at indicates the division between actively star-forming [ $\log_{10}(\text{SSFR}) > -11 \text{ yr}^{-1}$ ] and passive [ $\log_{10}(\text{SSFR}) < -11 \text{ yr}^{-1}$ ] galaxies determined in McGee et al. (2011). Bottom: same as the figure on the left except with a  $M_{\text{stellar}} > 1.4 \times 10^{10} M_{\odot}$  stellar mass cut applied.

dusty star-forming galaxies or edge-on discs with strong extinction, it seems unlikely that these galaxies would preferentially be found in groups with no substructure. An alternative explanation is that this population could be the ‘transition’ galaxies observed by Wolf (2009). These galaxies are still star forming but at a lower rate ( $\sim 4$  times) than the field and have more obscured star formation, resulting in weak optical signatures (i.e. not blue). A population

**Table 5.** Percentage of all galaxies in our sample within a given region of  $^{0.4}(g-r)$  versus SSFR space.

$^{0.4}(g-r)$ –SSFR region	Galaxies in groups with substructure	Galaxies in groups with no substructure	Field galaxies
Red and passive <sup>a</sup>	$28 \pm 6$	$42 \pm 5$	$17 \pm 1$
Red and active <sup>b</sup>	$3 \pm 3$	$11 \pm 4$	$4 \pm 1$
Blue and passive	$9 \pm 4$	$9 \pm 3$	$9 \pm 1$
Blue and active	$60 \pm 7$	$38 \pm 5$	$70 \pm 2$

<sup>a</sup>Passive denotes quiescent galaxies with  $\log_{10}(\text{SSFR}) < -11 \text{ yr}^{-1}$ .<sup>b</sup>Active refers to actively star-forming galaxies with  $\log_{10}(\text{SSFR}) > -11 \text{ yr}^{-1}$ .**Table 6.** Same as Table 5 except for galaxies above the stellar mass completeness limit of  $M_{\text{stellar}} > 1.4 \times 10^{10} M_{\odot}$ .

$^{0.4}(g-r)$ –SSFR region	Galaxies in groups with substructure	Galaxies in groups with no substructure	Field galaxies
Red and passive	$42 \pm 8$	$57 \pm 7$	$33 \pm 2$
Red and active	$4 \pm 5$	$14 \pm 6$	$8 \pm 1$
Blue and passive	$13 \pm 7$	$8 \pm 5$	$11 \pm 1$
Blue and active	$41 \pm 8$	$21 \pm 5$	$48 \pm 2$

of ‘transition’ galaxies could explain why there are more ‘red and active’ galaxies in groups with no substructure, as a relaxed system could contain galaxies that are being quenched but have not had their star formation completely cut-off.

### 5.3 Implications of the observed properties of groups with substructure

The field-like colour distribution of the groups with substructure and the fact that the substructure galaxies are found in the group outskirts may have implications for the nature of environmental effects in galaxy evolution. It is well known that the properties of galaxies depend, at least in some part, on their local environment (Postman & Geller 1984; Dressler et al. 1997). Galaxies that reside in dense environments, such as groups or clusters, generally experience some form of star formation attenuation due to processes such as ram pressure stripping (Gunn & Gott 1972; Abadi, Moore & Bower 1999; Quilis, Moore & Bower 2000), strangulation (Larson, Tinsley & Caldwell 1980; Balogh, Navarro & Morris 2000; Kawata & Mulchaey 2008) or mergers and interactions (Toomre & Toomre 1972; Brough et al. 2006). However, the precise details of the galaxy transformation process (i.e. exactly when and where quenching occurs, which mechanisms dominate in the different environments, etc.) are still unclear.

Our substructure analysis in the GEEC groups suggests that the identified local substructure in our sample, which in some cases appears to be infalling, do not feel any strong environmental effects from the host group. The observed colours and SSFRs of the galaxies in groups with substructure are significantly more blue, active and remarkably field-like when compared to galaxies in groups with no substructure. This suggests that an enhanced fraction of red galaxies – either/both via the suppression of star formation or/and dust obscuration – only happens in relaxed groups with no detected substructure. Thus, any environmental effects felt by infalling substructure galaxies do not likely occur until well inside the group potential.

Recent studies of star formation as a function of group- or cluster-centric radius have produced conflicting results with regard to the radius at which environmental effects become observable. Similar to our results, Wetzel, Tinker & Conroy (2011) conclude that galaxies

do not show suppressed star formation outside the virial radius. In contrast, von der Linden et al. (2010) state that suppressed star formation could be detected in SDSS clusters out to several virial radii. Such differences are likely sensitive to group or cluster-finding algorithms as well as membership assignment.

In a study of the effects of environment on the colours of galaxies in the SDSS survey, Wilman, Zibetti & Budavári (2010) take a different approach to classifying environment. Rather than using a catalogue derived from a group-finding algorithm, these authors parametrize the environment using non-overlapping annular measurements of density on independent scales, allowing for comparison of environmental effects at various radii. Based on their analysis, Wilman et al. (2010) concluded that the fraction of red galaxies correlated with local density only up to scales of  $\sim 1$  Mpc, which is similar to our results, as well as those of Wetzel et al. (2011). Though numerous and independent analyses of environmental effects on galaxy evolution seem to indicate that star formation truncation does not occur until galaxies are well inside the group/cluster potential, there are studies that suggest the contrary. Clearly, more work, both observational and theoretical, is needed to better understand when and where environmental effects become observable.

## 6 CONCLUSIONS

We have studied the DS test for substructure to determine the sensitivity and reliability of this test for group-sized systems. Using mock groups with and without substructure, generated using Monte Carlo methods, we find that the DS test can reliably be applied to groups with more than 20 members, if the probabilities, or  $P$ -value, method is used with a high c.l. of 99 or 95 per cent. We also find that for groups with  $10 \leq n_{\text{members}} < 20$ , the DS test cannot detect all of the substructure within a system, but it can be used to determine a reliable *lower limit* on the amount of substructure.

Of the 15 rich GEEC groups, with a velocity dispersion range of  $\sim 260$ – $950 \text{ km s}^{-1}$  and  $20 < n_{\text{members}} < 90$ , we find that four groups are identified as having significant substructure. Further analysis indicates that two of these systems, GEEC Groups 208 and 320, likely have gravitationally bound local substructure that lies on the group outskirts and could be accreting on to the system.

We then looked at various dynamical and galaxy properties to search for correlations with the presence of substructure. The main results of this analysis are the following.

(i) The majority of groups with detected substructure also have non-Gaussian velocity distributions.

(ii) The shape of a group's VDP correlates with the detection of substructure, where GEEC groups with substructure have rising profiles.

(iii) The  $^{0.4}(g-r)$  colour distributions of the groups with and without substructure are found to be significantly different, and the colour distribution for the galaxies in groups with substructure is similar to the field distribution.

(iv) Groups with substructure have a significantly higher fraction of blue galaxies, as do the galaxies within identified regions of localized substructure.

(v) Groups with substructure have a larger fraction of actively star-forming galaxies [ $\log_{10}(\text{SFFR}) > -11 \text{ yr}^{-1}$ ], when compared to groups with no identified substructure.

(vi) There is a measurable fraction of galaxies that populate the 'red and active' region of the  $^{0.4}(g-r)$ -SSFR space, and we find that this fraction is significantly higher in groups with no substructure for both the whole and stellar mass limited samples.

In conclusion, we find that a considerable fraction of intermediate-redshift galaxy groups contain significant substructure, which suggests that like massive clusters, groups grow hierarchically through the accretion of smaller structures. The field-like colour distribution and measured SSFRs of the galaxies in groups with substructure, combined with the location of the substructure, suggest that these galaxies are not experiencing any form of environment-related star formation quenching. To fully understand the results presented here within the context of galaxy evolution will require the use of sophisticated modelling. To this end, we plan to duplicate this analysis on a sample of semi-analytic groups obtained from GALFORM simulations (Bower et al. 2006) and compare these with our observational results. With this we hope to be able to better understand the nature of substructure identified by the DS test.

## ACKNOWLEDGMENTS

We would like to thank the CNOC2 team for the use of unpublished redshifts. AH, LCP and WEH would like to thank the National Science and Engineering Research Council of Canada (NSERC) for funding them.

## REFERENCES

Abadi M. G., Moore B., Bower R. G., 1999, *MNRAS*, 308, 947  
 Balogh M. L., Navarro J. F., Morris S. L., 2000, *ApJ*, 540, 113  
 Balogh M. L. et al., 2009, *MNRAS*, 398, 754  
 Balogh M. L. et al., 2011, *MNRAS*, 412, 2303  
 Barrena R., Boschin W., Girardi M., Spolaor M., 2007, *A&A*, 469, 861  
 Beers T. C., Geller M. J., Huchra J. P., 1982, *ApJ*, 257, 23  
 Bergond G., Zepf S. E., Romanowsky A. J., Sharples R. M., Rhode K. L., 2006, *A&A*, 448, 155  
 Berlind A. A. et al., 2006, *ApJS*, 167, 1  
 Bird C., 1994a, *ApJ*, 422, 480  
 Bird C. M., 1994b, *AJ*, 107, 1637  
 Böhringer H. et al., 2010, *A&A*, 514, A32  
 Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *MNRAS*, 370, 645

Brough S., Forbes D. A., Kilborn V. A., Couch W., 2006, *MNRAS*, 370, 1223  
 Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000  
 Burgett W. S. et al., 2004, *MNRAS*, 352, 605  
 Cameron E., 2011, *PASA*, 28, 128  
 Carlberg R. G., Yee H. K. C., Morris S. L., Lin H., Hall P. B., Patton D. R., Sawicki M., Shepherd C. W., 2001, *ApJ*, 552, 427  
 Charlot S., Fall S. M., 2000, *ApJ*, 539, 718  
 Colless M., Dunn A. M., 1996, *ApJ*, 458, 435  
 D'Agostino R., Stephens M., 1986, *Goodness-of-fit Techniques*. Marcel Dekker Inc., New York  
 Diemand J., Kuhlen M., Madau P., Zemp M., Moore B., Potter D., Stadel J., 2008, *Nat*, 454, 735  
 Dressler A., Shectman S. A., 1988, *AJ*, 95, 985  
 Dressler A. et al., 1997, *ApJ*, 490, 577  
 Efron B., 1982, *The Jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia  
 Eke V. R., Baugh C. M., Cole S., Frenk C. S., King H. M., Peacock J. A., 2005, *MNRAS*, 362, 1233  
 Finoguenov A. et al., 2009, *ApJ*, 704, 564  
 Firth P., Evstigneeva E. A., Jones J. B., Drinkwater M. J., Phillipps S., Gregg M. D., 2006, *MNRAS*, 372, 1856  
 Geller M. J., Huchra J. P., 1983, *ApJS*, 52, 61  
 Girardi M., Fadda D., Giuricin G., Mardirossian F., Mezzetti M., Biviano A., 1996, *ApJ*, 457, 61  
 Girardi M., Demarco R., Rosati P., Borgani S., 2005, *A&A*, 442, 29  
 Gunn J. E., Gott J. R., III, 1972, *ApJ*, 176, 1  
 Hou A., Parker L. C., Harris W. E., Wilman D. J., 2009, *ApJ*, 702, 1199  
 Iovino A. et al., 2010, *A&A*, 509, A40  
 Katz N., White S. D. M., 1993, *ApJ*, 412, 455  
 Kawata D., Mulchaey J. S., 2008, *ApJ*, 672, L103  
 Knebe A., Müller V., 2000, *A&A*, 354, 761  
 Lacey C., Cole S., 1993, *MNRAS*, 262, 627  
 Lara-López M. A., Bongiovanni A., Cepa J., Pérez García A. M., Sánchez-Portal M., Castañeda H. O., Fernández Lorenzo M., Pović M., 2010, *A&A*, 519, A31  
 Larson R. B., Tinsley B. M., Caldwell C. N., 1980, *ApJ*, 237, 692  
 McGee S. L., Balogh M. L., Wilman D. J., Bower R. G., Mulchaey J. S., Parker L. C., Oemler A., 2011, *MNRAS*, 413, 996  
 Mamon G. A., 2007, in Saviane I., Ivanov V. D., Borissova J., eds, *The Evolution of Galaxy Groups and of Galaxies Therein*. Springer-Verlag, Berlin, p. 203  
 Menci N., Fusco-Femiano R., 1996, *ApJ*, 472, 46  
 Moore B., Katz N., Lake G., 1996, *ApJ*, 457, 455  
 Nelson L., 1998, *J. Quality Technol.*, 30, 298  
 Peng Y.-j. et al., 2010, *ApJ*, 721, 193  
 Pinkney J., Roettiger K., Burns J. O., Bird C. M., 1996, *ApJS*, 104, 1  
 Postman M., Geller M. J., 1984, *ApJ*, 281, 95  
 Press W. H., Schechter P., 1974, *ApJ*, 187, 425  
 Quilis V., Moore B., Bower R., 2000, *Sci*, 288, 1617  
 Ramella M. et al., 2007, *A&A*, 470, 39  
 Ribeiro A. L. B., Lopes P. A. A., Trevisan M., 2010, *MNRAS*, 409, L124  
 Salim S. et al., 2007, *ApJS*, 173, 267  
 Silverman B. W., 1986, *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London  
 Solanes J. M., Salvador-Solé E., González-Casado G., 1999, *A&A*, 343, 733  
 Springel V. et al., 2005, *Nat*, 435, 629  
 Springel V. et al., 2008, *MNRAS*, 391, 1685  
 Summers F. J., Davis M., Evrard A. E., 1995, *ApJ*, 454, 1  
 Taylor J. E., Babul A., 2004, *MNRAS*, 348, 811  
 Toomre A., Toomre J., 1972, *ApJ*, 178, 623  
 Tyler K. D. et al., 2011, *ApJ*, 738, 56  
 von der Linden A., Wild V., Kauffmann G., White S. D. M., Weinmann S., 2010, *MNRAS*, 404, 1231  
 Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., 2006, *MNRAS*, 366, 2  
 West M. J., Bothun G. D., 1990, *ApJ*, 350, 36

Wetzel A. R., Tinker J. L., Conroy C., 2011, ArXiv e-prints  
 Wilman D. J., Balogh M. L., Bower R. G., Mulchaey J. S., Oemler A.,  
 Carlberg R. G., Morris S. L., Whitaker R. J., 2005, MNRAS, 358, 71  
 Wilman D. J. et al., 2008, ApJ, 680, 1009  
 Wilman D. J., Oemler A., Mulchaey J. S., McGee S. L., Balogh M. L.,  
 Bower R. G., 2009, ApJ, 692, 298  
 Wilman D. J., Zibetti S., Budavári T., 2010, MNRAS, 406, 1701  
 Wolf C. et al., 2009, MNRAS, 393, 1302  
 Yang X., Mo H. J., van den Bosch F. C., Pasquali A., Li C., Barden M.,  
 2007, ApJ, 671, 153  
 Yee H. K. C. et al., 2000, ApJS, 129, 475  
 Zabludoff A. I., Mulchaey J. S., 1998a, ApJ, 496, 39  
 Zabludoff A. I., Mulchaey J. S., 1998b, ApJ, 498, L5

## APPENDIX A: FALSE NEGATIVE RATES IN MORE DETAIL

In Section 2.2.3, we presented the main results of the effects of changing various input parameters in our mock groups. Here we present tables detailing the specific false negative rates obtained and also discuss each free parameter in detail. As discussed in Section 2.2.3, we determine the sensitivity of the DS test to each of the free parameters in our mock groups ( $\sigma_{\text{position}}$ ,  $\epsilon_{\text{position}}$ ,  $\sigma_{\text{redshift}}$ ,  $\epsilon_{\text{redshift}}$ ,  $n_{\text{sub}}$  and  $\sigma_{\text{host}}$ ) by beginning with a ‘base’ mock group (Table 1), which has a false negative rate of zero per cent. We then change only one parameter at a time to ensure that any change in the false negative rate can be directly attributed to the altered free parameter.

### A1 Angular size of the substructure ( $\sigma_{\text{position}}$ )

In Table A1, we present the false negative rates (listed as a percentage) for mock groups with  $n_{\text{member}}$  values of 10, 15, 20 and 50 as a function of projected angular size on the sky ( $\sigma_{\text{position}}$ ). The first row in the table indicates the results for our ‘base’ groups, which have  $P$ -values of either 0 or 1 per cent. We then increase the value of  $\sigma_{\text{position}}$ , and from Table A1, we can see that the DS test reliably identifies substructure with a projected dispersion of up to 0.1 Mpc for groups with 10 members, and as large as 0.1 Mpc for groups with more than 15 members. If we increase  $\sigma_{\text{position}}$  to 0.2 Mpc, we find that for groups with roughly 20 members or less, the false negative rates increases dramatically, but still remain very low (1 per cent) for richer groups with 50 members. The general conclusion from Table A1 is that the DS test is not very sensitive to the size of the substructure and that even for small groups it can identify real substructure that is relatively large ( $\sim 0.1$  Mpc).

**Table A1.** False negative rates: dependency on the angular size of the input substructure ( $\sigma_{\text{position}}$ ).

$\sigma_{\text{position}}$ (Mpc)	$n_{\text{members}} = 10$	$n_{\text{members}} = 15$	$n_{\text{members}} = 20$	$n_{\text{members}} = 50$
0.01 <sup>a</sup>	1 <sup>b</sup>	0	0	0
0.05	3	1	0	0
0.09	4	1	3	0
0.1	11	3	2	0
0.2	48	26	20	1
0.5	83	74	73	27

<sup>a</sup>These mock groups have the following input parameters:  $\epsilon_{\text{position}} = 0.5$  Mpc,  $\sigma_{\text{redshift}} = 100 \text{ km s}^{-1}$ ,  $\epsilon_{\text{redshift}} = 1300 \text{ km s}^{-1}$  and  $\sigma_{\text{host}}$  values listed in Table 1. Only the  $\sigma_{\text{position}}$  parameter is varied for these trials.

<sup>b</sup>Values quoted are the false negative rates, given as a percentage, obtained for 100 trials with each set of inputs, using 100 000 MC shuffles.

**Table A2.** False negative rates: dependency on the location of the input substructure in position space ( $\epsilon_{\text{position}}$ ).

$\epsilon_{\text{position}}$ (Mpc)	$n_{\text{members}} = 10$	$n_{\text{members}} = 15$	$n_{\text{members}} = 20$	$n_{\text{members}} = 50$
0.001 <sup>a</sup>	7 <sup>b</sup>	2	0	0
0.01	11	2	3	0
0.1	6	0	0	0
0.2	0	1	0	0
0.3	2	0	0	0
0.4	1	0	0	0
0.5	3	0	0	0
1.0	0	1	0	0

<sup>a</sup>These mock groups have the following input parameters:  $\sigma_{\text{position}} = 0.01$  Mpc,  $\sigma_{\text{redshift}} = 100 \text{ km s}^{-1}$ ,  $\epsilon_{\text{redshift}} = 1300 \text{ km s}^{-1}$  and  $\sigma_{\text{host}}$  values listed in Table 1. Only the  $\epsilon_{\text{position}}$  parameter is varied for these trials.

<sup>b</sup>Values quoted are the false negative rates, given as a percentage, obtained for 100 trials with each set of inputs, using 100 000 MC shuffles.

### A2 Location of substructure in position space ( $\epsilon_{\text{position}}$ )

In Table A2, we list the false negative rates for mock groups with  $n_{\text{member}}$  values of 10, 15, 20 and 50 as a function of the projected radial distance of the substructure with respect to the group centroid ( $\epsilon_{\text{position}}$ ). It is clear from this table that the DS test is quite insensitive to  $\epsilon_{\text{position}}$ . In other words, the test can reliably identify substructure that is ‘close’ to the projected group centre, and easily detects structure that lies on the group outskirts.

### A3 Velocity dispersion of the input substructure ( $\sigma_{\text{redshift}}$ )

In Table A3, we list the false negative rates for mock groups with  $n_{\text{member}}$  values of 10, 15, 20 and 50 as a function of the velocity dispersion of the input substructure ( $\sigma_{\text{redshift}}$ ). From Table A3 it is evident that the DS test is also insensitive to this parameter and can reliably detect substructure with a wide range of velocity dispersions for all values of  $n_{\text{members}}$ . Only for groups with a fewer than 20 members and a very large dispersion value of  $450 \text{ km s}^{-1}$  do the false negative rates go above 5 per cent.

**Table A3.** False negative rates: dependency on the velocity dispersion of the input substructure ( $\sigma_{\text{redshift}}$ ).

$\sigma_{\text{redshift}}$ ( $\text{km s}^{-1}$ )	$n_{\text{members}} = 10$	$n_{\text{members}} = 15$	$n_{\text{members}} = 20$	$n_{\text{members}} = 50$
50 <sup>a</sup>	0 <sup>b</sup>	0	0	0
100	0	0	1	0
150	1	1	0	0
200	0	0	0	0
250	4	2	0	0
300	3	4	0	0
350	3	2	0	0
400	3	4	1	0
450	13	13	3	0

<sup>a</sup>These mock groups have the following input parameters:  $\sigma_{\text{position}} = 0.01$  Mpc,  $\epsilon_{\text{position}} = 0.5$  Mpc,  $\epsilon_{\text{redshift}} = 1300 \text{ km s}^{-1}$  and  $\sigma_{\text{host}}$  values listed in Table 1. Only the  $\sigma_{\text{redshift}}$  parameter is varied for these trials.

<sup>b</sup>Values quoted are the false negative rates, given as a percentage, obtained for 100 trials with each set of inputs, using 100 000 MC shuffles.



**A4 Location of substructure along the line of sight ( $\epsilon_{\text{redshift}}$ )**

In Table A4, we list the false negative rates for mock groups with  $n_{\text{member}}$  values of 10, 15, 20 and 50 as a function of the location of the substructure along the line of sight ( $\epsilon_{\text{redshift}}$ ). This parameter is taken to be a displacement (in  $\text{km s}^{-1}$ ) of the peak in the velocity distribution of the substructure with respect to the peak of the host's distribution. Unlike the previously discussed parameters, the DS test appears to be extremely sensitive to the location of the substructure along the line of sight, or more specifically the separation between the main groups' velocity distribution and the substructure velocity distribution.

The first set of entries in Table A4 are the false negative rates from mock groups with  $\sigma_{\text{host}}$  values set to the average dispersion of observed GEEC groups with similar group membership (Table 1). It is clear that if the peaks of the host and substructure velocity distributions are close ( $<300 \text{ km s}^{-1}$ ), then the DS test cannot always identify real substructure. For groups with fewer than  $\sim 20$  members, the peaks must be at least  $900 \text{ km s}^{-1}$  apart in order for the false negative rates to fall below 5 per cent.

We also find not only that is the DS test sensitive to the  $\epsilon_{\text{redshift}}$  parameter, but also that the level of sensitivity is dependent on the number of members in the host group. This is best seen by looking

**Table A4.** False negative rates: dependency on the location of the input substructure along the line of sight (i.e. redshift space) ( $\epsilon_{\text{redshift}}$ ).

$\epsilon_{\text{redshift}}$ ( $\text{km s}^{-1}$ )	$n_{\text{members}} = 10$	$n_{\text{members}} = 15$	$n_{\text{members}} = 20$	$n_{\text{members}} = 50$
100 <sup>a</sup>	91 <sup>b</sup>	94	80	43
200	85	81	74	39
300	62	71	72	21
400	53	49	54	7
500	32	42	39	0
600	25	25	18	0
700	10	19	7	0
800	7	9	9	0
900	3	5	4	0
1000	3	2	4	0
1100	2	0	0	0
1200	1	1	0	0
1300	1	0	0	0
100 <sup>c</sup>	84	94	87	51
200	83	87	67	43
300	80	81	69	21
400	63	66	48	4
500	43	49	34	4
600	37	43	21	0
700	35	26	13	0
800	21	22	3	0
900	17	8	1	0
1000	10	10	0	0
1100	5	15	0	0
1200	5	2	0	0
1300	2	5	0	0

<sup>a</sup>These mock groups have the following input parameters:  $\sigma_{\text{position}} = 0.01 \text{ Mpc}$ ,  $\epsilon_{\text{position}} = 0.5 \text{ Mpc}$ ,  $\epsilon_{\text{redshift}} = 1300 \text{ km s}^{-1}$  and  $\sigma_{\text{host}}$  values listed in Table 1. Only the  $\sigma_{\text{redshift}}$  parameter is varied for these trials.

<sup>b</sup>Values quoted are the false negative rates, given as a percentage, obtained for 100 trials with each set of inputs, using 100 000 MC shuffles.

<sup>c</sup>These mock groups have the following input parameters:  $\sigma_{\text{position}} = 0.01 \text{ Mpc}$ ,  $\epsilon_{\text{position}} = 0.5 \text{ Mpc}$ ,  $\epsilon_{\text{redshift}} = 1300 \text{ km s}^{-1}$  and a constant  $\sigma_{\text{host}}$  value of  $500 \text{ km s}^{-1}$  for all values of  $n_{\text{members}}$ . Again, only the  $\sigma_{\text{redshift}}$  parameter is varied for these trials.

**Table A5.** False negative rates: dependency on the number of members in the input substructure ( $n_{\text{sub}}$ ).

$n_{\text{sub}}$	$n_{\text{members}} = 10$	$n_{\text{members}} = 15$	$n_{\text{members}} = 20$	$n_{\text{members}} = 50$
3 <sup>a</sup>	43 <sup>b</sup>	39	54	51
4	1	2	9	16
5		0	0	3
6				0
7				0
8				0
9				0

<sup>a</sup>These mock groups have the following input parameters:  $\sigma_{\text{position}} = 0.01 \text{ Mpc}$ ,  $\epsilon_{\text{position}} = 0.5 \text{ Mpc}$ ,  $\sigma_{\text{redshift}} = 100 \text{ km s}^{-1}$ ,  $\epsilon_{\text{redshift}} = 1300 \text{ km s}^{-1}$  and  $\sigma_{\text{host}}$  values given in Table 1. For these simulations we only change the number of members in the substructure ( $n_{\text{sub}}$ ).

<sup>b</sup>Values quoted are the false negative rates, given as a percentage, obtained for 100 trials with each set of inputs, using 100 000 MC shuffles. A null entry indicates that no trials were run with the associated  $n_{\text{sub}}$  value.

at the false negative rates listed in the second set of values listed in Table A4, where we use a constant value of  $\sigma_{\text{host}} = 500 \text{ km s}^{-1}$  for all values of  $n_{\text{members}}$ . From this section of the table we see that when the velocity distribution of the substructure is located at  $1\sigma_{\text{host}}$ , the false negative rates are 43, 49, 34 and 4 per cent for mock groups with 10, 15, 20 and 50 members. At  $2\sigma_{\text{host}}$ , the rates are 10 per cent for groups with 10 and 15 members and 0 per cent for groups with 20 and 50 members. This result indicates that for groups with fewer member galaxies, the velocity distributions of the host and substructure must be very distinct in order for the DS test to detect substructure. Alternatively, groups with more than 20 members can have substructure galaxies with a velocity distribution embedded within the host distribution and still be identified by the test.

**A5 Number of members in the input substructure ( $n_{\text{sub}}$ )**

In Table A5 we list the false negative rates for mock groups with  $n_{\text{member}}$  values of 10, 15, 20 and 50 as a function of the number of members in the input substructure ( $n_{\text{sub}}$ ) (see Section 2.2.3 for discussion).

**Table A6.** False negative rates: dependency on the velocity dispersion of the host group ( $\sigma_{\text{host}}$ ).

$\sigma_{\text{host}}$ ( $\text{km s}^{-1}$ )	$n_{\text{members}} = 10$	$n_{\text{members}} = 15$	$n_{\text{members}} = 20$	$n_{\text{members}} = 50$
100 <sup>a</sup>	0 <sup>b</sup>	0	0	0
200	1	0	0	0
300	0	0	0	0
400	1	0	0	0
500	2	2	0	0
600	10	3	1	0
700	9	10	3	0
800	21	20	3	0
900	22	28	3	0

<sup>a</sup>These mock groups have the following input parameters:  $\sigma_{\text{position}} = 0.01 \text{ Mpc}$ ,  $\epsilon_{\text{position}} = 0.5 \text{ Mpc}$ ,  $\sigma_{\text{redshift}} = 100 \text{ km s}^{-1}$  and  $\epsilon_{\text{redshift}} = 1300 \text{ km s}^{-1}$ . For these simulations we keep the values of the input substructure constant, but change the velocity dispersion of the host group ( $\sigma_{\text{host}}$ ).

<sup>b</sup>Values quoted are the false negative rates, given as a percentage, obtained for 100 trials with each set of inputs, using 100 000 MC shuffles.



**A6 Velocity dispersion of the host group ( $\sigma_{\text{host}}$ )**

In addition to the above tests, we also check to see if changing the velocity dispersion, and therefore mass, of the host group ( $\sigma_{\text{host}}$ ) affects the observed false negative rates. We present the results in Table A6, where we see that for groups with more than 20 members the dispersion of the host group does not significantly increase the

rate of false negatives. However, for systems with fewer members and larger values of  $\sigma_{\text{host}}$  ( $>700 \text{ km s}^{-1}$ ), the DS test is more likely to miss true substructure. Fortunately, observed groups with 10 or so members do not generally have such high dispersion values.

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.